04-04-07

AF/1652

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Fernandez, et al.	Examiner: Fronda, C.
Serial No.: 10/003,021	Group Art Unit: 1652
Filed: November 14, 2001	Docket No. IVGN 276.1 CON
For: **Libraries of Expressible Gene Sequences**	**TRANSMITTAL LETTER**

Mail Stop Appeal Brief-Patents
Commissioner for Patents
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Transmitted herewith are the following documents in the above-identified application.

- [X] Brief on Appeal Under 37 C.F.R. § 41.37
- [X] Exhibit 1; Dubensky, et al., U.S. Pat. No. 6,342,372
- [X] Exhibit 2; Guan, et al., EP Pat. No. 0286239B1
- [X] Exhibit 3; Gregoire, et al., (*J. Biol Chem.*, 1996, Dec 20; 271(51):32951-9)
- [X] Express Mail Return Receipt Postcard; EV 655818524 US

CERTIFICATE OF EXPRESS MAILING

NUMBER: EV 655818524 US
DATE OF DEPOSIT: April 2, 2007

I hereby certify that this paper or fee is being deposited with the U.S. Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Mail Stop Appeal Brief – Patents, Commissioner for Patents, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450

By: Melody A. Ouellet

In the event that the Patent Office determines that these documents were not timely filed within the one-month deadline, the Commissioner is hereby authorized to charge the **Deposit Account 50-3994** for any fees due in connection with the filing of this document according to § 1.17(a)(1).

Respectfully submitted,

Date: April 2, 2007

/Natalie A. Davis/
Natalie A. Davis
Reg. No. 53,849
Agent for Appellants

Invitrogen Corporation
1600 Faraday Avenue
Carlsbad, CA 92008
Phone: (760) 268-7469

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Confirmation No.: 2174
FERNANDEZ, et al.,	Art Unit: 1652
Appl. No. 10/003,021	Examiner: Fronda, C.
Filed: November 14, 2001	Atty. Docket: IVGN 276.1 CON
For: **Libraries of Expressible Gene Sequences**	

Brief on Appeal Under 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

A Notice of Appeal from the final rejection of claims 41-43 and 45-58 was filed on February 1, 2007. Appellants hereby file this Appeal Brief, together with the required brief filing fee under § 41.20(b)(2) of $500.00.

It is not believed that extensions of time are required beyond those that may otherwise be provided for in documents accompanying this paper. However, if additional extensions of time are necessary to prevent abandonment of this application, then such extensions of time are hereby petitioned under 37 C.F.R. § 1.136(a), and any fees required therefor are hereby authorized to be charged to our Deposit Account No. 50-3994.

Table of Contents

I. Real Party In Interest

The real party in interest in this appeal is Invitrogen Corporation.

II. Related Appeals and Interferences

No other prior or pending appeals, interferences or judicial proceedings are known to the Appellants, the Appellants' legal representative, or assignee which may be related to, or directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. Status of Claims

Claims 41-43 and 45-58 are pending in the application.

Claims 1-40, 44, and 59-66 have been canceled.

Claims 41-43 and 45-58 are rejected.

IV. Status of Amendments

No amendments were filed subsequent to the final rejection.

V. Summary of Claimed Subject Matter

Claims 41 and 58 are the independent claims involved in this Appeal. The invention defined by claim 41 relates generally to isolated expression vectors. The

expression vector comprises a 5'-CACC sequence linked immediately 5' to a start codon of an open reading frame (ORF). The ORF is linked in-frame to a polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. Support for claim 41 can be found throughout the specification, for example, at page 2, lines 1-6 and lines 13-27 through page 4, line 22; page 7, line 26 through page 8, line 27; page 10, lines 7-17; page 12, line 26 through page 13, line 4; Example 1 at pages 8-21; Example 2 at page 78, line 3 through page 79, line 19; Table 1 at pages 21-78; Table 2 at pages 79-146; and Example 3 at pages 147-148.

Claim 58 relates generally to libraries of expression vectors. The libraries comprise a plurality of expression vectors, where each vector comprises a 5'-CACC sequence linked immediately 5' to a start codon of an open reading frame (ORF). The ORF is linked in-frame to the polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. The ORF of an expression vector in the plurality may be the same or different from open reading frames of other expression vectors in the plurality. Support for claim 58 can be found throughout the specification, for example, at page 2, lines 1-6 and 13-27 through page 4, line 22; page 7, line 26 through page 8, line 27; page 10, lines 7-17; page 12, line 26 through page 13, line 4; Example 1 at pages 8-21; Example 2 at page 78, line 3 through page 79, line 19; Table 1 at pages 21-78; Table 2 at pages 79-146; and Example 3 at pages 147-148.

VI. Grounds of Rejection to be Reviewed on Appeal

Claims 41-43 and 45-58 stand rejected under 35 U.S.C. 103(a), as being unpatentably obvious over Dubensky, et al., (U.S. Pat. No. 6,342,372), in view of Guan, et al., (EP Pat. No. 0286239B1) and Gregoire, et al., (*J. Biol Chem.,* 1996, Dec 20; 271(51):32951-9).

VII. Argument

A. Legal Standard for Obviousness

Establishing prima facie obviousness requires a showing that each claim element is taught or suggested by the prior art. See *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). Absent a showing of such motivation and suggestion, prima facie obviousness is not established. See *In re Fine*, 837 F.2d 1071 (Fed Cir 1988). The Court of Appeals for the Federal Circuit has indicated that:

> The PTO has the burden under section 103 to establish a prima facie case of obviousness...It can satisfy this burden only by showing some objective teaching in the prior art or that knowledge generally available to one of ordinary skill in the art would lead that individual to combine the relevant teachings of the references. Id at 1074.

To meet its burden, the PTO "cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention." Id. at 1075. The Court of Appeals for the Federal Circuit has held numerous times that such hindsight analysis is impermissible. Instead, the PTO must show suggestions, explicit or otherwise, that would compel one of ordinary skill to combine the cited

references in order to make and use the claimed invention. *See, e.g., Interconnect Planning Corp. v. Feil*, 774 f.2d 1132, 1143 (Fed. Cir. 1985).

Further, the PTO must consider prior art references in their entirety, *i.e.* as a whole, including portions that teach away from the claimed invention. *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984). The Court of Appeals for the Federal Circuit has instructed that "references that teach away cannot serve to create a prima facie case of obviousness" (*In re Gurley*, 27 F.3d 551, 553 (Fed. Cir. 1994)), and that an "applicant may rebut a prima facie case of obviousness by showing that the prior art teaches away from the claimed invention in any material respect" (*In re Geisler*, 116 F.3d 1465, 1469 (Fed. Cir. 1997)).

1. The Cited References

a. The Dubensky Reference

The Dubensky reference discloses eukaryotic vector systems for the production of recombinant proteins, where the vectors include a CACC sequence linked 5' to the ATG start codon of a nucleic acid encoding a heterologous polypeptide. The disclosed vectors are configured in a "bicistronic heterologous configuration" specifically designed to prevent the expression of fusion proteins. This is because heterologous genes in Dubensky's bicistronic vectors are separated by a stop codon so that the encoded proteins are expressed separately. See column 90, paragraphs 2 and 3.

The Dubensky reference does not disclose expression vectors having an open reading frame (ORF) linked in-frame to a polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. Rather heterologous proteins encoded by Dubensky's bicistronic vectors are expressed separately and are not associated as a fusion protein.

b. The Guan Reference

The Guan reference discloses vectors that encode fusion proteins; specifically polypeptides linked to maltose binding protein. See column 1, lines 1-14; column 3, line 53 through column 4 line 3; and column 5, line 55 through column 6, line 18. The disclosed fusion protein is purified by affinity chromatography using the maltose binding protein. See column 1, lines 13-18 and column 12, lines 40-52.

The Guan reference does not disclose an isolated expression vector comprising the sequence 5'-CACC linked immediately 5' to a start codon of an open reading frame or an expression library comprising such vectors.

c. The Gregoire Reference

The Gregoire reference discloses a vector that encodes a fusion protein; specifically a recombinant form of the horse allergen Equ c1 protein linked to a polyhistidine tail. The disclosed fusion protein is purified by affinity chromatography using the polyhistidine tail. See the abstract; page 32951, column 2, third paragraph; figure 1 on page 32952; and figure 2 on page 32954 and column 1, third paragraph.

The Gregoire reference does not disclose an isolated expression vector comprising the sequence 5'-CACC linked immediately 5' to a start codon of an open reading frame or a fusion protein or an expression library comprising such vectors.

2. The Examiner's Position

The Examiner argues that the methods of claims 41-43 and 45-58 are obvious over Dubensky in view of Guan and Gregoire.

The Examiner states that the Dubensky reference teaches an oligonucleotide primer comprising a CACC sequence linked 5' to the ATG start codon of a nucleic acid encoding a heterologous polypeptide. The Examiner recognizes that Dubensky does not disclose expression vectors having an open reading frame (ORF) linked in-frame to a polynucleotide encoding a heterologous peptide thereby encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. The Examiner offers the Guan and Gregorie references to cure this deficiency.

The Guan and Gregorie references are offered to address the shortcomings of Dubensky: an open reading frame (ORF) that is linked in-frame to a polynucleotide encoding a heterologous peptide, thus encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. Specifically, the Guan reference is offered for its disclosure of polypeptides linked to a maltose binding protein. The Gregorie is offered to specifically address Dubenskys' shortcoming of an affinity purification tag, which is required by claims 45 and 46. The Gregorie reference is said to disclose a recombinant protein with a polyhistidine tail.

As to motivation to combine these references, the Examiner simply states that it would have been obvious to a skilled artisan to modify Dubenskys' bicistronic expression vectors in the manner disclosed by Guan, to produce fusion proteins suitable for purification. The Examiner further states that it would have been obvious to further modify Dubensky's polynucleotide to encode a polyhistidine tail as disclosed in Gregorie. Specifically, the Examiner states:

> It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the polynucleotide comprising the CACC sequence (SEQ ID No. 69) linked to the 5' start codon ATG of nucleic acids encoding heterologous peptides taught by Dubensky Jr. et al. such that the DNA encoding the MBP and DNA encoding a peptide that can be recognized and cut by a protease as taught by Guan et al. is linked to the polynucleotide taught by Dubensky Jr. et al. Alternatively, the polynucleotide taught by Dubensky Jr. et al. is modified to have a DNA encoding a polyhistidine tail as taught by Gregorie et al. See Office Action dated August 1, 2006 at page 3.

3. The Appellant's Position

Claims 41-43 and 45-58 are not obvious over the cited references.

Claims 41-43 and 45 are drawn to isolated expression vector, comprising (1) the sequence 5'-CACC linked immediately 5' to a start codon of (2) an open reading frame (ORF) linked in-frame to a polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising the ORF-encoded polypeptide and the heterologous peptide. Claim 58 is drawn to libraries of such isolated expression vectors.

The Examiner appears to be relying on hindsight in combining the Dubensky, Guan and Gregoire references in arriving at an obviousness determination. Worse, in so doing, the Examiner has failed to consider the cited references in their entirety, including the parts that teach away from the claimed invention.

The Examiner has ignored the fact that Dubensky's vectors are configured in a "bicistronic heterologous configuration" specifically designed to prevent the expression of fusion proteins. Applicants have brought this fact to the Examiner's attention in response to the Office Actions dated October 19, 2005 and August 1, 2006. Dubensky's bicistronic vectors include a stop codon between heterologous genes, and therefore are suitable only for the expression of single peptides. See column 90, paragraphs 2 and 3. Dubensky's bicistronic vectors are specifically designed not to (and cannot) encode fusion proteins as required by the present claims.

The Examiner's selective reading of Dubensky is contrary to the proscription of the Court of Appeals for the Federal Circuit – references that teach away cannot serve to create a prima facie case of obviousness. *In re Gurley*, 27 F.3d 551, 553 (Fed. Cir. 1994). Moreover, ignoring a key aspect of the Dubensky disclosure to focus only on claim elements that are disclosed, and combining those with the remaining elements found in other references is not appropriate – it is a clear example of impermissible hindsight analysis.

The Examiner has not provided any "suggestions, explicit or otherwise, that would compel one of ordinary skill to combine the cited references in order to make and use the claimed invention," as required by the Court of Appeals for the Federal Circuit.

See *In re Fine* at 1071. Applicants therefore respectfully request that this rejection under 35 U.S.C. § 103 be withdrawn.

D. Conclusion

In view of the forgoing discussion, Appellants respectfully submit that the subject matter defined by claims 41-43 and 45-58 are patentable over the cited art. Appellants therefore respectfully request that the Board reverses the Examiner's final rejection of the pending claims and remand this application for issue.

Respectfully submitted,

Date: April 2, 2007

/Natalie A. Davis/
Natalie A. Davis
Reg. No. 53,849
Agent for Appellants

Invitrogen Corp.
1600 Faraday Ave.
Carlsbad, CA 92008
(760) 268-7469

VIII. Claims Appendix

41. An isolated expression vector, comprising the sequence 5'-CACC linked immediately 5' to a start codon of an open reading frame (ORF), wherein the ORF is linked in-frame to a polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising a polypeptide encoded by the ORF and the heterologous peptide.

42. The expression vector of claim 41, wherein the ORF encodes a full length polypeptide.

43. The expression vector of claim 41, wherein the ORF lacks a stop codon.

45. The expression vector of claim 41, wherein the heterologous peptide comprises an affinity purification tag or an epitope tag.

46. The expression vector of claim 41, wherein the heterologous peptide comprises a polyhistidine tag, a chitin binding domain, glutathione-S-transferase, biotin, or a V5 epitope.

47. The expression vector of claim 41, further comprising a polynucleotide encoding an endopeptidase recognition sequence linked in-frame between the ORF and the polynucleotide encoding the heterologous peptide.

48. The expression vector of claim 41, which is a eukaryotic expression vector or a prokaryotic expression vector.

49. The expression vector of claim 41, which is suitable for prokaryotic expression and eukaryotic expression.

50. The expression vector of claim 41, which is suitable for expression in bacteria cells, fungi, insect cells, yeast cells, plant cells, or mammalian cells.

51. The expression vector of claim 41, further comprising a promoter, an enhancer sequence, a selection marker sequence, an origin of replication, an epitope-tag encoding sequence, an affinity purification-tag encoding sequence, or a combination thereof.

52. The expression vector of claim 51, wherein the promoter is a constitutive promoter or an inducible promoter.

53. The expression vector of claim 52, wherein the constitutive promoter is a T7 promoter, a β-lactamase gene promoter, a bacteriophage λ int promoter; a chloramphenicol acetyl transferase gene promoter, an SV40 promoter, an RSV promoter or a CMV promoter.

54. The expression vector of claim 52, wherein the inducible promoter is a trp promoter, a recA promoter, a lacZ promoter, a lacI promoter, an araC promoter, an I-amylase promoter, a metallothionein I gene promoter, a herpesvirus TK promoter, an SV40 early promoter, a yeast gal1 gene promoter, an EF1 promoter, or an ecdysone-responsive promoter.

55. The expression vector of claim 51, wherein the selection marker confers resistance to ampicillin, tetracycline, kanamycin, bleomycin, streptomycin, hygromycin, neomycin, or Zeocin™ antibiotic.

56. The expression vector of claim 51, wherein the selection marker is a hisD gene sequence or a URA3 sequence.

57. The expression vector of claim 51, wherein the origin of replication (ori) is an *Escherichia coli* oriC ori, a yeast 2μ ori, a yeast ARS ori, and sf1 ori, or an SV40 ori.

58. A library of expression vectors, comprising a plurality of expression vectors, wherein each expression vector comprises the sequence 5'-CACC linked immediately 5' to a start codon of an open reading frame (ORF), wherein said ORF is linked in-frame to a polynucleotide encoding a heterologous peptide, thereby encoding a fusion protein comprising a polypeptide encoded by the ORF and the heterologous peptide, and wherein an ORF of an expression vector in the plurality is the same or different from open reading frames of other expression vectors in the plurality.

IX. Evidence Appendix

Exhibit	Title of Exhibit	Location in Record
Exhibit 1	Dubensky, et al., U.S. Pat. No. 6,342,372	Cited by Examiner in Office Action dated August 1, 2006
Exhibit 2	Guan, et al., EP Pat. No. 0286239B1	Cited by Examiner in Office Action dated August 1, 2006
Exhibit 3	Gregoire, et al., (*J. Biol Chem.*, 1996, Dec 20; 271(51):32951-9)	Cited by Examiner in Office Action dated August 1, 2006

X. Related Proceedings Appendix

None.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 286 239 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
13.03.1996 Bulletin 1996/11

(51) Int Cl.[6]: **C12N 15/00**, C12N 11/00, C12P 21/02

(21) Application number: **88302039.8**

(22) Date of filing: **09.03.1988**

(54) **Production and purification of a protein fused to a binding protein**

Herstellung und Reinigung eines Proteins, das mit einem Bindungsprotein fusioniert ist

Production et purification d'une protéine fusionnée d'une protéine de liage

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(30) Priority: **10.03.1987 US 24053**

(43) Date of publication of application:
12.10.1988 Bulletin 1988/41

(73) Proprietors:
- **NEW ENGLAND BIOLABS, INC.**
  Beverly Massachusetts 01915 (US)
- **Temple University of the Commonwealth System of Higher Education**
  Philadelphia PA 19122 (US)

(72) Inventors:
- **Guan, Chudi**
  Beverly Massachusetts 01915 (US)
- **Inouye, Hiroshi**
  Deceased (US)

(74) Representative: **Bass, John Henton et al**
REDDIE & GROSE
16 Theobalds Road
London WC1X 8PL (GB)

(56) References cited:
EP-A- 0 157 235 **EP-A- 0 195 680**
EP-A- 0 244 147

- **CHEMICAL ABSTRACTS, vol. 97, no. 17, 25th October 1982, page abstract no. 140527e, Columbus, Ohio, US; K. ITO.**
- **GENE, vol. 29, July 1984, pages 27-31, Amsterdam, NL; A. ULLMANN: "One-step purification of hybird proteins which have beta-galactosidase activity"**
- **Ito et al. (1982), Journal of Biological Chemistry Vol. 257 pp. 9895-9897.**
- **Bassford et al. (1979), Journal of Bacteriology, Vol. 139, pp. 19-31.**

EP 0 286 239 B1

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

The present invention relates to a process of producing and/or purifying virtually any hybrid polypeptide or fused protein molecule employing recombinant DNA techniques. More specifically, a DNA fragment coding for a protein molecule, e.g. a polypeptide or portion thereof, is fused to a DNA fragment coding for a binding protein such as the gene coding for the maltose binding protein. The fused DNA is inserted into a cloning vector and an appropriate host transformed. Upon expression, a hybrid polypeptide or fused protein molecule is produced which can be purified by contacting the hybrid polypeptide with a ligand or substrate to which the binding protein has specific affinity, e.g. by affinity chromatography. The hybrid polypeptide so purified may in certain instances be useful in its hybrid form, or it may be cleaved to obtain the protein molecule itself by, for example, linking the DNA fragments coding for the protein molecule and binding protein with a DNA segment which codes for a peptide which is recognized and cut by a proteolytic enzyme. The present invention also relates to certain vectors useful in practicing the above process as well as to a bioreactor and methods employing the bound hybrid polypeptide, e.g. where the bound fused polypeptide is contacted and reacted with a susbstrate which interacts with the bound protein molecule to produce a desired result.

Recently developed techniques have made it possible to employ microorganisms, capable of rapid and abundant growth, for the synthesis of commercially useful proteins and peptides. These techniques make it possible to genetically endow a suitable microorganism with the ability to synthesize a protein or peptide normally made by another organism. In brief, DNA fragments coding for the protein are ligated into a cloning vector such as a plasmid. An appropriate host is transformed with the cloning vector and the transformed host is identified, isolated and cultivated to promote expression of the desired protein. Proteins so produced are then isolated from the culture medium for purification.

Many purification techniques have been employed to harvest the proteins produced by recombinant DNA techniques. Such techniques generally include segregation of the desired protein based on its distinguishing molecular properties, e.g. by dialysis, density-gradient centrifugation and liquid column chromatography. Such techniques are not universally applicable and often result in consumption of the purification materials which may have considerably more value than the protein being purified, particularly where substantial quantities of highly purified protein are desired.

Other procedures have been developed to purify proteins based on solubility characteristics of the protein. For example, isoelectric precipitation has been employed to purify proteins since the solubility of proteins varies as a function of pH. Similarly, solvent fractionation of proteins is a technique whereby the solubility of a protein varies as a function of the dielectric constant of the medium. Solvent fractionation, while giving good yields often causes denaturation of the protein molecule. Neither isoelectric precipitation nor solvent fractionation are useful in obtaining highly purified protein. Such techniques are typically employed in tandem with other procedures.

Proteins have also been separated based on their ionic properties by e.g. electrophoresis, ion-exchange chromatography, etc. Such electrophoretic techniques, however, have been used as analytical tools and are not practical as a means for purifying proteins on a large scale. Moreover, high purity and yield of the protein obtainable by such techniques is rarely achieved in a single step.

Affinity chromatography has also been employed in the purification of biopolymers such as proteins. Affinity chromatography involves a selective adsorbent which is placed in contact with a solution containing several kinds of substances including the desired species to be purified. For example, when used in protein purification protocols, affinity chromatography generally involves the use of a ligand which specifically binds to the protein to be purified. In general, the ligand is coupled or attached to a support or matrix and the coupled ligand contacted with a solution containing the impure protein. The non-binding species are removed by washing and the desired protein recovered by eluting with a specific desorbing agent. While affinity chromatography produces a relatively high level of purified protein, this technique requires significant amounts of the protein-specific ligand employed for purification. Moreover, the ligand will be different for each and every protein to be purified which necessarily entails a time-consuming and laborious regime. In addition, it has been found that specific ligands do not exist for all types of protein molecules, such as certain enzymes. As a result, affinity chromatography has not been successfully employed as a universal isolation purification technique for protein molecules.

One proposed attempt to universalize affinity chromatography to all proteins is described in European Patent Application 0,150,126 (Hopp). Disclosed is the preparation of a hybrid molecule produced by recombinant DNA techniques employing gene fusion. One gene codes for the desired protein to be purified while the other codes for an identification or marker peptide. The marker peptide contains a highly antigenic N-terminal portion to which antibodies are made and a linking portion to connect the marker peptide to the protein to be purified. The linking portion of the marker peptide is cleavable at a specific amino acid residue adjacent the protein molecule to be purified by use of a specific proteolytic agent. The fused or hybrid protein is isolated by constructing an affinity column with immobilized antibody specific to the antigenic portion of the marker peptide. The antibody binds to the fused protein which can thereafter be liberated from the column by a desorbing agent. The marker

peptide may then be cleaved from the desired protein molecule with a proteolytic agent.

While purportedly overcoming some of the problems described above for protein purification protocols, Hopp requires substantial amounts of antibodies specific for the antigenic portion of the marker peptide. Moreover, the quantity of desorbing agent (in this case, a small peptide) required to compete off the target protein is substantial as well as a significant cost factor. Also, the desorbing agent must be purified away from the target protein. Thus, scale up for this system would not be practical. Furthermore, regeneration of the chromatographic column may be extremely difficult due to the destabilizing conditions employed to wash out the column after *use,* which may, in fact destroy the column. Others have suggested the use of low affinity antibody columns. However, low affinity columns often result in non-specific binding and would require significant cost for any large scale purification.

Thus, there is a continuing need for techniques which enable large scale purification of proteins produced through recombinant DNA processes without the above described problems. It would be particularly advantageous to provide an affinity purification process which utilizes an abundant and inexpensive ligand to which the fused protein would bind and an equally abundant and inexpensive desorbing agent.

SUMMARY OF THE INVENTION

In accordance with the present invention there is provided a method for producing and highly purifying virtually any protein molecule generated by recombinant DNA techniques in a single affinity chromatography step. The method comprises:

(a) constructing a DNA expression vector which expresses a hybrid polypeptide in a transformed host cell, the hybrid polypeptide comprising the target protein molecule and a non-enzymatic biologically functional sugar binding protein, having a specific affinity for a substrate which binds to the non-enzymatic biologically functional sugar binding protein; and

(b) introducing the expression vector into an appropriate host cell and expressing the hybrid polypeptide;

(c) contacting the hybrid polypeptide produced by the transformed cell with the substrate to which the non-enzymatic biologically functional sugar binding protein binds; and

(d) recovering the target protein molecule.

The hybrid polypeptide or fused protein is produced by recombinant DNA techniques. The hybrid polypeptide can be isolated and purified directly, e.g. from the crude cellular extract or culture medium, simply by contacting the extract containing the hybrid polypeptide with a substrate to which the binding protein has specific affinity, e.g. using affinity chromatography. The bound hybrid polypeptide can easily be liberated from the column in a highly purified form with a desorbing agent which selectively desorbs the bound non-enzymatic sugar binding protein. While the target protein may be useful in its hybrid form, in certain preferred embodiments, it may be desirable to separate or cleave the non-enzymatic sugar binding protein away from the target protein. This may be accomplished in a variety of ways. For example, a DNA fragment coding for a predetermined peptide, e.g. a linking sequence, may be employed to link the DNA fragments coding for the binding and target proteins. The predetermined peptide is preferably one which is recognized and cleaved by a proteolytic agent such that it cuts the hybrid polypeptide at or near the target protein without interfering with the biological activity of the target protein. The linking sequence, in addition to providing a convenient proteolytic cleavage site, may also serve as a polylinker, i.e. by providing multiple DNA restriction sites to facilitate fusion of the DNA fragments coding for the target and binding proteins, and/or as a spacer which separates the target and binding protein which, for example, allows access by the proteolytic agent to cleave the fused polypeptide.

The preferred affinity column useful in practicing the present invention, in general, comprises a column containing immobilized ligand or substrate to which the binding protein has a specific affinity. As will be appreciated by the skilled artisan, the specific affinity of a binding protein for a given substrate will depend both on the particular binding protein employed as well as the substrate used in the column. In general, the substrate used in the column should bind substantially all of the particular binding protein without binding other proteins to which it is exposed. In certain instances, however, depending on the particular application (e.g. whether the column is used to purify the protein molecule or as a bioreactor for reacting the protein molecule with a substance with which it interacts to produce a desired result), a substrate may be used which only binds a portion of the binding protein present. In addition, the particular substrate employed should permit selective desorbtion of the bound binding protein with a suitable desorbing agent.

It will be appreciated that the column thus prepared can be used to isolate and purify virtually any protein which, by recombinant DNA techniques is linked to the binding, protein to form a hybrid polypeptide. The hybrid polypeptide can be released from the column with a suitable desorbing agent and/or cleaved with a proteolytic agent to separate the target protein from the binding protein. Alternatively, in accordance with another embodiment of the present invention, the bound hybrid polypeptide may be used as a bioreactor for reacting, for example, the biologically active portion of the protein molecule

(which may be an enzyme, restriction endonuclease, etc.) with a substrate which interacts with the target protein. For example, if the target protein is an enzyme, the affinity column can serve as a means for immobilizing that enzyme, i.e. by the binding protein portion of the hybrid polypeptide being bound to the column. The substrate upon which the enzyme acts is thereafter passed through the column to achieve the desired result.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1 and 2 illustrate the construction of the maltose binding protein fusion cloning vector pCG150.

Figure 3 illustrates the DNA sequence of the polylinker region of the cloning vector pCG150.

Figure 4 illustrates the constuction of the mal E - Lac Z gene fusion plasmid pCG325.

Figure 5 illustrates elution profile of the protein resulting from affinity chromatography of a crude extract of SF1362/pCG325 containing the mal E - Lac Z fusion.

Figure 6 illustrates the activity profile of the protein resulting from affinity chromatography of a crude extract of SF1362/pCG325 containing the mal E - Lac Z fusion .

Figure 7 illustrates the SDS polyacrylamide gel electrophoresis of the product of the mal E - Lac Z fusion.

Figure 8 illustrates the native polyacrylamide gel electrophoresis of the product of the mal E - Lac Z fusion.

Figures 9 and 10 illustrate the construction of the mal E - Pst I restriction endonuclease gene fusion plasmid pCG410.

Figure 11 illustrates the SDS polyacrilamide gel electrophoresis of the product of the mal E - Pst I fusion.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a novel approach for producing and purifying a target protein molecule comprising:

(a) constructing a DNA expression vector which expresses a hybrid polypeptide in a transformed host cell, the hybrid polypeptide comprising the target protein molecule and a non-enzymatic biologically functional sugar binding protein, having a specific affinity for a substrate which binds to the non-enzymatic biologically functional sugar binding protein; and

(b) introducing the expression vector into an appropriate host cell and expressing the hybrid polypeptide;

(c) contacting the hybrid polypeptide produced by the transformed cell with the substrate to which the non-enzymatic biologically functional sugar binding protein binds; and

(d) recovering the target protein molecule.

The protein molecule is produced by constructing a DNA expression vector containing fused genes comprising a gene encoding the protein molecule and a gene coding for a non-enzymatic biologically functional sugar binding protein or portion thereof which has a specific affinity for a ligand or substrate and expressing the fusion in an appropriate host. The substrate is used as the matrix in an isolation/purification protocol, e.g. an affinity column, to recover the expressed product of the fused genes, i.e. the hybrid polypeptide. A DNA fragment which codes for a predetermined polypeptide can be used, e.g. flanking the gene coding for the binding protein,- in order to adjust the reading frame for the desired gene fusion and/or to introduce into the hybrid polypeptide a peptide sequence which is recognized and cleaved by a proteolytic agent which enables separation of the protein molecule from the binding protein where desired. As noted above, the bound hybrid polypeptide may also be used as a bioreactor for reacting the biologically active portion of the protein molecule with a substrate which interacts with the protein molecule.

The methods described herein by which DNA coding for a hybrid polypeptide is preferably cloned, expressed and purified include the following steps:

I. Preparation of Fusion Vector.

A) The DNA encoding for the desired binding protein is purified.

B) The DNA is inserted into a cloning vector such as pBR322 and the mixture is used to transform an appropriate host such as E. coli.

C) The transformants are selected, such as with antibiotic selection or other phenotypic selection.

D) The plasmid DNA is prepared from the selected transformants.

E) The binding activity domain of the protein is determined and convenient restriction endonuclease sites are identified by mapping or created by standard genetic engineering methods.

II. Insertion of DNA Coding for the Protein Molecule into the Fusion Vector.

A) The protein molecule gene is cloned by standard genetic engineering methods.

B) The protein molecule gene is characterized, e.g. by restriction mapping.

C) A DNA restriction fragment which encodes the protein molecule is prepared.

D) The protein molecule DNA fragment is inserted in the binding protein fusion vector so that an in-frame protein fusion is formed between the the DNA fragment coding for the binding protein and the DNA fragment coding for the protein molecule.

E) The vector containing this hybrid DNA molecule

is introduced into an appropriate host.

III. Expression and Purification of the Hybrid Polypeptide.

A) The host cell containing the fusion vector is cultured.
B) Expression of the fused gene is induced by conventional techniques.
C) A cell extract containing the expressed fused polypeptide is prepared.
D) The hybrid polypeptide is separated from other cell constitutants using an affinity column having as a matrix a substance to which the non-enzymatic biologically functional sugar binding protein part of the hybrid polypeptide has a specific affinity.
E) The bound purified hybrid polypeptide can be recovered and/or utilized by the following methods:

(1) if the protein molecule's biological activity is maintained in its hybrid or fused configuration it may recovered from the column by eluting with a desorbing agent and used directly after elution in its hybrid form;
(2) the protein molecule may be separated from the non-enzymatic biologically functional sugar binding protein either before or after elution from the column by proteolytic or chemical cleavage; and
(3) the column may be used as a bioreactor with the fusion protein immobilized on the column, e.g. by contacting and reacting the bound fusion protein with a substrate which interacts with the biologically active portion of the protein molecule.

Binding Protein

Non-enzymatic biologically functional sugar binding proteins which may be employed in accordance with the present invention include the sugar (e.g. mono-, di- or polysaccharide) binding proteins such as maltose or arabinose binding protein.

The preferred sugar binding protein for practicing the present invention is the maltose binding protein.

The product of the mal E Gene of E. coli, i.e. maltose binding protein (MBP) is a periplasmic osmotically shockable protein. MBP exhibits specific binding affinity with maltose and maltodextrins. Macromolecular alpha (1-4) linked glucans are also bound with high affinities. Ferenci, T. and Klotz, U. Escherichia Coli. FEBS Letters, Vol. 94, No. 2. pp. 213-217 (1978), the disclosure of which is hereby incorporated by reference. The dissociation constants are around 1um. Kellermann et al., Coli Eur. J. Biochem. 47. 139-149 (1974), the disclosure of which is hereby incorporated by reference. MBP is usually considered to exist as a monomer although it can exist as a dimer. Maltose induces the conversion of the dimer to the monomer. Gilbert, Biochemical and Biophysical Research Communications (1982) Vol. 105, No. 2, pp. 476-481, the disclosure of which is hereby incorporated by reference. MBP is a secreted protein which is synthesized in cytoplasm as a precursor with a 26 amino acid N-terminal signal peptide. Dupley, et al. J. Biol. Chem. Vol. 259 pp. 10606-10613 (1984), the disclosure of which is hereby incorporated by reference. During translocation across the cytoplasmic membrane the signal peptide is removed and the mature MBP is released into the periplasmic space. Mature MPB contains 370 amino acids corresponding to a molecular weight of 40,661 dalton(Dupley, et al., supra). MBP is made in large quantity in an induced culture (2-4x10^4 monomers per cell). It has been determined that MBP and at least four other proteins make up the maltose transport system of E. coli. Shuman, J. Biol. Chem. 257: 5455-5461 (1982), the disclosure of which is hereby incorporated by reference. Besides being an essential component of the maltose transport system, MBP is also the specified chemoreceptor of the bacterium for maltose and maltodextrins. The Mal E gene has been cloned and sequenced. Dupley, et al., supra.

Linking Sequence

A DNA fragment coding for a predetermined peptide may be employed to link the DNA fragments coding for the binding protein and protein molecule. The predetermined peptide is preferably one which is recognized and cleaved by a proteolytic agent such that it cuts the hybrid polypeptide at or near the protein molecule without interfering with the biological activity of the protein molecule. One such DNA fragment coding for a predetermined polypeptide is described in Nagai et al., Nature, Vol. 309, pp. 810-812 (1984), the disclosure of which is hereby incorporated by reference. This DNA fragment has the oligonucleotide sequence: ATCGAGGGTAGG and codes for the polypeptide Ile-Glu-Gly-Arg. This polypeptide is cleaved at the carboxy side of the arginine residue using blood coagulation factor Xa. As noted above the linking sequence, in addition to providing a convenient cut site, may also serve as a polylinker, i.e. by providing multiple restriction sites to facilitate fusion of the DNA fragments coding for the target and binding proteins, and/or as a spacing means which separates the target and binding protein which, for example, allows access by the proteolytic agent to cleave the hybrid polypeptide.

Protein Molecule

The present invention may be beneficially employed to produce substantially any prokaryotic or eukaryotic, simple or conjugated protein that can be expressed by a vector in a transformed host cell. Such proteins include enzymes including endonucleases, methylases, oxi-

doreductases, transferases, hydrolases, lyases, isomerases or ligases.

The present invention also contemplates the production of storage proteins, such as ferritin or ovalbumin or transport proteins, such as hemoglobin, serum albumin or ceruloplasmin. Also included are the types of proteins that function in contractile and motile systems, for instance, actin and myosin.

The present invention also contemplates the production of antigens or antigenic determinants which can be used in the preparation of vaccines or diagnostic reagents.

The present invention also contemplates the production of proteins that serve a protective or defense function, such as the blood proteins thrombin and fibrinogen. Other protective proteins include the binding proteins, such as antibodies or immunoglobulins that bind to and thus neutralize antigens.

The protein produced by the present invention also may encompass various hormones such as Human Growth Hormone, somatostatin, prolactin, estrone, progesterone, melanocyte, thyrotropin, calcitonin, gonadotropin and insulin. Other such hormones include those that that have been identified as being involved in the immune system, such as interleukin 1, intereukin 2, colony stimulating factor, macrophage-activating factor and interferon.

The present invention is also applicable to the production of toxic proteins, such as ricin from castor bean or grossypin from cotton linseed.

Proteins that serve as structural elements may also be produced by the present invention; such proteins include the fibrous proteins collagen, elastin and alpha-keratin. Other structural proteins include glyco-proteins, virus-proteins and muco-proteins.

In addition to the above-noted naturally occuring proteins, the present invention may be employed to produce synthetic proteins defined generally as any sequences of amino acids not occurring in nature.

Genes coding for the various types of protein molecules identified above may be obtained from a variety of prokaryotic or eukaryotic sources, such as plant or animal cells or bacteria cells. The genes can be isolated from the chromosome material of these cells or from plasmids of prokaryotic cells by employing standard, well-known techniques. A variety of naturally occuring and synthetic plasmids having genes encoding many different protein molecules are now commercially available from a variety of sources. The desired DNA also can be produced from mRNA by using the enzyme reverse transciptase. This enzyme permits the synthesis of DNA from an RNA template.

Preparation of DNA Fusion and Expression Vectors

Various procedures and materials for preparing recombinant vectors; transforming host cells with the vectors; replicating the vector and expressing polypeptides and proteins; are known by the skilled artisan and are discussed generally in Maniatis et al., Molecular Cloning: A Laboratory Manual, CSH 1982, the disclosure of which is hereby incorporated by reference.

In practicing the present invention, various cloning vectors may be utilized. Although the preferred vector is a plasmid, the skilled artisan will appreciate that the vector may be a phage. If cloning takes place in mammalian or plant cells, viruses can also be used as vectors. If a plasmid is employed, it may be obtained from a natural source or artificially synthesized. The particular plasmid chosen should be compatible with the particular cells serving as the host, whether a bacteria such as E. coli, yeast, or other unicellular microorganism. The plasmid should also have the proper origin of replication (replicon) for the particular host cell chosen. In addition, the capacity of the vector must be sufficient to accommodate the fusion coding for both the protein molecule of interest and the binding protein.

Another requirement for a plasmid cloning vector is the existence of restriction enzymes to cleave the plasmid for subsequent ligation with the foreign genes without causing inactivation of the replicon while providing suitable ligatable termini that are complementary to the termini of the foreign genes being inserted. To this end, it would be helpful for the plasmid to have single substrate sites for a large number of restriction endonucleases.

Moreover, the plasmid should have a phenotypic property that will enable the transformed host cells to be readily identified and separated from cells which do not undergo transformation. Such phenotypic selection genes can include genes providing resistance to a growth inhibiting substance, such as an antibiotic. Plasmids are now widely available that include genes resistant to various antibiotics, such as tetracycline, streptomycin, sulfa drugs, and ampicillin. When host cells are grown in a medium containing one of these antibiotics, only transformants having the appropriate resistant gene will survive.

If E. coli is employed as the host cell, a preferred plasmid for performing the present invention is pCG150. A partial restriction endonuclease cleavage map of this plasmid is shown in Figure 2. An alternative plasmid for high level expression in E. coli is pCG806.

To prepare the chosen plasmid for ligation, preferably, it is digested with a restriction endonuclease to produce a linear segment(s) in which the two DNA strands are cleaved at closely adjacent sites to produce cohesive termini ("sticky ends") bearing 5'-phosphate and 3'-hydroxyl groups, thereby facilitating ligation with the foreign genes. For the plasmids identified above, restriction endonucleases will produce this result.

Certain restriction enzymes (Pvu II, Bal I) may result in the formation of blunt ends. The blunt ends of the plasmid can be joined to the foreign genes with T4 DNA ligase. The methods and materials for achieving efficient cleavage and ligation are well known in the art.

Prior to being joined with the selected cloning vector, it is desirable that the foreign genes coding for the binding protein and the protein molecule be first joined together. Ideally, the gene coding for the protein molecule molecule is treated with the same restriction endonuclease used to cleave the plasmid vector so that the appropriate termini of the gene will be compatible with the corresponding termini of the plasmid. This gene also may be treated with a second, different restriction endonuclease to prepare its opposite terminus for ligation with the binding protein gene.

The cointegrate genes are next ligated to the linearized plasmid fragment in a solution with DNA ligase. After incubation, the recircularized plasmid having the correct orientation of the cointegrate genes are identified by standard techniques, such as by gel electrophoresis.

Transformation of Recombinant DNA Plasmid.

The recombinant DNA plasmids, as prepared above, are used for the transformation of host cells. Although the host cell may be any appropriate prokaryotic or eukaryotic cell, preferably it is well-defined bacteria, such as E. coli or yeast strain. Both such hosts are readily transformed and capable of rapid growth in fermentation cultures. In place of E. coli, other unicellular microorganisms can be employed, for instance fungae and algae. In addition, other forms of bacteria such as salmonella or pneumococcus may be substituted for E. coli. Whatever host is chosen, it should be one that has the necessary biochemical pathways for phenotypic expression and other functions for proper expression of the hybrid polypeptide. The techniques for transforming recombinant plasmids in E. coli strains are widely known. A typical protocol is set forth in Maniatus et al. supra.

In transformation protocols, only a small portion of the host cells are actually transformed, due to limited plasmid uptake by the cells. Thus, before transformants are isolated, the host cells used in the transformation protocol typically are multiplied in an appropriate medium. The cells that actually have been transformed can be identified by placing the original culture on agar plates containing a suitable growth medium containing the phenotypic identifier, such as an antibiotic. Only those cells that have the proper resistance gene will survive. Cells from the colonies that survive can be lysed and then the plasmid isolated from the lysate. The plasmid thus isolated can be characterized, e.g. by digestion with restriction endonucleases and subsequent gel electrophoresis or by other standard methods.

Once transformed cells are identified, they can be multiplied by established techniques, such as by fermentation. In addition, the recovered cloned recombinant plasmids can be used to transform other strains of bacteria or other types of host cells for large scale replication and expression of the fused protein.

Purification of the Fused Protein

The hybrid polypeptide expressed by the transformed host cell are preferably separated from all other cellular constitutents and growth media by an affinity chromatography process. The column matrix is simply any substrate for which the binding protein has specific affinity. For example, when the binding protein is MBP the column matrix may be crosslinked amylose. Crosslinked amylose prepared by an epichlorohydrin protocol satisfies the substrate specificity of MBP and provides a rapid one step chromatographic purification of MBP from osmotic-shock fluids, Ferenci, T. et al., supra, whole cell extracts or culture media.

An extract from the transformed host cell is contacted with the column to isolate the hybrid polypeptide. The hybrid polypepetide may thereafter be eluted from the column, for example, by adding a dilute solution of a desorbing agent which displaces the hybrid polypeptide.

Separation of the Protein Molecule from the Hybrid Polypeptide

The hybrid polypeptide purified from the above affinity column may be cleaved by sequence specific proteases such as a factor Xa or by discrete chemical cleavage such as cyanogen bromide.

The following examples are given to additionally illustrate embodiments of the present invention as it is preferred to practice. It should be understood that these examples are illustrative, and that the invention is not to be considered as restricted thereto except as indicated in the appended claims.

EXAMPLE I

Example I describes cloning, expression and purification of B-galactosidase as a product of the mal E - Lac Z gene fusion.

Preparation of the Binding Protein Fusion Vector

Plasmid pPL-5A is the source for the Mal E encoding DNA fragment which is prepared by first creating a deletion derivative of pPL-5A which moves the Mal E promoter and signal sequence. This plasmid is pCG810. The gene encoding Mal E is then resected from pCG810 and inserted into M13mp18 to produce recombinant phage pCG580, which has added multiple cloning sites to facilitate insertion of protein molecule encoding DNA. The Male E gene now carrying the additional cloning site is resected from pCG580 and inserted into pUC18 in order to create additional cloning sites as well as pick up a selective antibiotic resistance gene. The resulting plasmid is the protein fusion vector pCG150 which contains the Mal E gene and additional cloning sites and which is used in the construction of the vector which also contains the DNA coding for the desired protein molecule, infra.

A sample of pCG150 has been deposited with the American Type Culture Collection under ATCC accession No. 67345. The construction of plasmid pCG150 is illustrated in Figs. 1 and 2.

According to the published Mal E gene sequence of E. coli there are five Taq I recognition sites in the gene. One is located at base number 83-86 (Dupley, et al. supra) corresponding to the second and third codon of mature maltose binding protein (MBP) coding sequence. A kanamycin resistance determinant fragment flanked by polylinkers was inserted into this Taq I site. The resulting plasmid was pPL-5A.

5-10 ug of pPL-5A plasmid DNA and 10 units of EcoRI restriction enzyme in 100ul of EcoRI digestion buffer was incubated for 2 hours at 37°C. 20ul of DNA gel loading buffer (0.25% bromophenol blue, 40mM EDTA, pH 8.0, 30% glycerol) were added and mixed. The digested sample was applied to 1% low gelling temperature agarose gel (Seaplaque). Gel electrophoresis was performed at low current (20mA) for 4 hours. TEA gel electrophoresis buffer (40mM Tris-acetate, pH 8.0. 2mM EDTA) was used. The gel was stained with TEA buffer containing ethidium bromide 0.5 ug/ml for 30 minutes at room temperature. Three DNA bands were visualized on the gel by U.V. irradiation. The largest fragment was cut out of the gel and placed in a 1.5 ml microfuge tube. The tube was incubated for 5 minutes in a 65°C water bath. The melted gel (about 100ul) was extracted with an equal volume of phenol and phenol/chloroform and chloroform as described by Maniatis et al, supra, at page 170, the disclosure of which is hereby incorporated by referernce. The aqueous phase was saved and 1/10 volume of 3N sodium-acetate pH 5.5 was added and mixed. 2.5 volumes of ethanol was added. The ethanol precipitate mixture was placed in -70°C freezer for 20 minutes (or in -20°C freezer overnight), then centrifuged for 15 minute in a microfuge at 4°C. The supernatant was discarded and the pellet was rinsed with 0.5 ml of 70% ethanol twice. The tube was left open at room temperature to eliminate any remaining ethanol. The DNA pellet was dissolved in 19 ul of water followed by adding 4 ul of 6x ligation buffer (300mM Tris-HCl pH 7.4, 60mM Mg Cl_2, 60mM dithiothreitol, 6 mM ATP, 600ug BSA) and 1ul of T4 DNA ligase (10 units) and incubated at 16°C overnight. The ligation solution was used to transform competent cells of E. coli strain SF 1362. The competent cells were made and the transformation was performed as described by T.J. Silhavy et al., in Experiments with Gene Fusions, CSH pp. 169-170 (1984), the disclosure of which is hereby incorporated by reference. After heat shock the transformation mixture was incubated with 5 ml LB medium for 45 minutes at 37°C. The cells were collected by centrifugation for 5 minutes at 3000 r.p.m. and resuspended in 0.5 ml of LB medium. 0.05-0.2 ml of the cells were spread on LB plates containing ampicillin 100 ug/ml. After overnight incubation at 37°C a total of about 1000 transformants were obtained. 16 transformants were purified on the same plates. Plasmid DNA minipreparations from the purified transformants were performed as described by Silhavy et al., supra. Restriction enzyme analysis on the plasmid DNAs was also performed. One plasmid was chosen, pCG810, in which the kanamycin resistance determenent sequence and the malE promotor and signal sequence regions had been deleted and the single EcoR,I BglII, BssHII and NcoI cutting sites remained.

10-20 ug of plasmid pCG810 DNA prepared and purified by the BND cellulose procedure described by Gamper et al., DNA, Vol. 4, No.2 (1985), the disclosure of which is hereby incorporated by reference, and 20 units of Hinf I restriction enzyme in 100ul of Hinf I digestion buffer (recommended by N.E.B.) were incubated for 2 hours at 37°C then extracted with phenol and chloroform and precipitated with ethanol as described above. The DNA was dissolved in 50 ul of the filling in reaction buffer (50mM Tris. pH 7.4. 10mM $MgCl_2$, 1mM dithiothreitol, 0.1mM dATP, 0.1mM dCTP, 0.1mM dGTP and 0.1mM dTTP containing 5 units of DNA polymerase I large fragment and incubated for 20 minutes at room temperature. 50 ul of TE buffer (10mM Tris. pH 8.0, 1mM EDTA) were added and extracted with phenol and chloroform and the aqueous phase precipitated with ethanol. The DNA was cleaved with EcoRI restriction enzyme in 100 ul of EcoRI digestion buffer followed by ethanol precipitation. The DNA was redissolved in 50 ul of TE followed by 10 ul of DNA gel loading buffer and applied to 1% of low gelling temperature agarose gel. The gel electrophoresis and DNA extraction from gel were as described above. The 1.1 kb EcoRI-Hinf I fragment which contained almost the entire MBP coding sequence was purified and dissolved in 10 ul of DNA buffer (10mM Tris pH 8.0, 0.1mM EDTA), stored at -20°C.

5 ug of M13mp18 double stranded DNA (Yanisch-Perron et al., Gene: 33, pp.103-119 at 104, (1985)), the disclosure of which is hereby incorporated by reference, and 10 units of SmaI restriction enzyme in 50 ul of SmaI digestion buffer were incubated for 30 minutes at 37°C followed by phenol extraction and ethanol precipitation as described above. The digested DNA was then dissolved in 50 ul of EcoRI digestion buffer containing 10 units EcoRI restriction enzyme and incubated for 1 hour, then extracted with phenol and chloroform, precipitated with ethanol as described above. The DNA pellet was dissolved in 10 ul of DNA buffer.

Two DNA preparations, the 1.1 kb EcoRI-HinfI fragment and the EcoRI and SmaI digested M13mp18 vector, were pooled and ligation was performed as described above. The ligation solution was used to transform JM101 or 71-18 competent cells (Yanisch-Peron et al., supra). The transformation was done as described above. After the heat shock the cells were mixed with JM101 or 71-18 exponentially growing cells and melted soft agar keeped at 47°C and plated on LB plates containing XG and IPTG described by J. Messing in NIH Publication No. 79-99, Vol. 2, (1979) at 43-48, the disclosure of which is hereby incorporated by reference.

About 500 to 1000 plaques appeared on the plate; 60% were white, 40% blue. About 100 white plaques were picked up with sterile pasteur pipets and added to 5 ml culture tubes containing 2 ml early log phase culture of JM101 or 71-18. The tubes were incubated for 5-6 hours at 37°C with shaking. The phage containing supernatants were seperated from the cells by transfering 1 ml each of culture into a microfuge tube and centrifugation for 10 minutes with microfuge at room temperature. 20 ul of supernatant were withdrawn and mixed with 1 ul of 2% S.D.S. and 4 ul of DNA gel loading buffer. Samples were electrophoresed through 0.8% agarose gel in 4xTAE buffer overnight. The recombinant phages were identified by slower migration through the gel as compared with single stranded DNA of phage M13mp18. Double stranded DNAs were made from the recombinant phages and restriction enzyme analyses were carried out. One recombinant phage pCG580 was chosen which had the Mal E gene sequence insertion in the same direction as Lac Z gene on M13mp18, in which the EcoRI cutting site was regenerated. The BamHI-XbaI-SalI-PstI-SphI-HindIII polylinker remained. BglII, BssHII and NcoI cutting sites were introduced in by the insertion of the malE sequence.

5 ug of pCG580 double stranded DNA purified with BND cellulose was cleaved with EcoRI restriction enzyme followed by blunting the cohesive ends with DNA polymerase I large fragment as described above. The DNA was religated and used to transform JM101 or 71-18. Only less than 5% of transformants were blue. It seemed that the filling in EcoRI cutting site created an in-frame TAA codon which could not be suppressed by Sup E carried by JM101. The small portion of blue transformants could be explained by a base deletion from the cohesive ends during the DNA manipulation and indicated the inserted Mal E sequence was in the same reading frame with down stream Lac Z sequence since no detectable DNA deletion was found for the plasmids made from the blue transformants by restriction enzyme analyses.

10-20 ug of double stranded pCG580 DNA purified with BND cellulose was cleaved with EcoRI. After phenol extraction and ethanol precipitation the DNA pellet was dissolved in 100 ul of mung bean exonuclease buffer containing about 5 units mung bean exonuclease and incubated for 20 minutes at 37°C followed by phenol extraction and ethanol precipitation. The blunted DNA was then cleaved with Hind III restriction enzyme in 50 ul of Hind III digestion buffer. This sample was electrophoresed through 1% of low gelling temperature agarose gel. The 1.1 kb DNA fragment containing MBP coding sequence tailed with polylinker was purified from the gel as described above. The purified DNA fragment was stored in 10 ul of DNA buffer at -20°c.

10 ug of pUC-18 plasmid DNA and 20 units of BamH1 restriction enzyme in 100 ul of BamH1 digestion buffer were incubated for 1-2 hours at 37°C. After phenol extraction and ethanol precipitation the digested DNA was treated with mung bean exonuclease to blunt the cohesive ends as described above. After phenol extraction and ethanol precipitation the DNA was dissolved in 10 ul of DNA buffer.

Two DNA preparations, the 1.1 kb fragment from pCG580 and the BamHI cleaved pUC-18, were pooled and 4 ul of 6x ligation buffer and 1 ul of T_4 ligase (5-10 units) were added and mixed. The ligase solution was incubated overnight at 16°C followed by incubation for 4 hours at room temperature and used to transform JM103 or 71-18. Transformants were selected on LB plates containing ampicillin 100 ug/ml. Recombinant plasmids were identified by the size of DNA with the toothpick assay as described by Shinmick et al., Nucl. Acids Res. Vol. 2, p. 1911, the disclosure of which is hereby incorporated by reference. About 12 recombinant plasmids were scored and three produced blue color on LB ampicillin plates in the presence XG and IPTG. One was chosen as plasmid pCG150. 5 ug of pCG150 plasmid DNA purified with BND cellulose was cleaved with EcoRI restriction enzyme followed by blunting the cohesive ends with large fragment DNA polymerase I, then ligated with T_4 Ligase. When this DNA was used to transform JM101 or 71-18, more that 95% of transformants were white in presence of XG and IPTG. This indicated no translation restarted in the downstream Mal E gene region.

The Mal E gene joint regions on plasmid pCG150 were sequenced and the results presented in Fig 3.

The Mal E - B-galactosidase fusion protein plasmid pCG325 illustrated in Fig. 4 was constructed as follows. Plasmid pMLB1034 was constructed by Silhavy et al, supra. This plasmid contains the Lac Z gene coding for B-galactosidase without the promotor or first 8 codons of the protein and a polylinker containing EcoRI, SmaI and BamHI restriction sites. 5 ug of pMLB1034 was cleaved with EcoRI restriction enzyme followed by blunting the cohesive ends with DNA polymerase large fragment, then cleaved with BamHI. After phenol extraction and ethanol precipitation the DNA was dissolved in 10 ul of DNA buffer and stored at -20°C.

5 ug of pCG150 DNA was cleaved with BamHI and PVUII restriction enzymes, extracted with phenol chloroform, precipitated with ethanol. The DNA was dissolved in 10 ul of DNA buffer. Two pCG150 and PMLB1034 DNA preparations were pooled and ligated as described above. The ligation solution was used to transform competent cells made from an E. coli straim MC4100 Silhavy, T.J., et al, supra and spread on LB plates containing ampicillin 100 ug/ml, XG 20 ug/ml. After overnight incubation several hundred transformants appeared on plates, 20-30% of them were blue. About 24 blue transformants were purified and used to isolate plasmid DNAs usingh the rapid isolation method described by Silhavy, supra. Restriction enzyme analyses were performed on these plasmid DNAs.

One recombinant, plasmid pCG325, was chosen and characterized. This plasmid contained the 1.3kb Mal E gene sequence from pCG150 which had been inserted

in the EcoRI-BamHI site of pMLB1034.

Affinity Chromatography

A double deletion (ΔLacΔmalB) strain E. coli (SF1362) habouring pCG325 was grown to late log phase in rich medium containing ampicillin 100 ug/ml. Cells were harvested by centrifugation with a Beckman centrifuge for 15 minutes at 5000 r.p.m. at 4°C. 5 gms of harvested cells were washed with 100 ml of 10mM TRIS. pH 7.2 at 4°C, then resuspended in 50 ml of the same buffer. Cells were broken by sonication at 4°C. Cell debris was separated by centrifugation with a Beckman centrifuge for 30 minutes at 16000 r.p.m. The supernatant was dialysed against 1 L of the same buffer for 3-4 hours at 4°C. A sample was applied onto a 3 x 5 cm cross-linked amylose column prepared as described by Ferenci et al., supra at pp. 459-463.

After the major 280 mu absorbant peak passed through at about 20-30 ml the column was extensively washed with 10-20 column volume of 10mM Tris pH 7.2. The column was eluted with 10mM Tris, pH 7.2, containing 10mM maltose. Both O.D 280mu and B-galactosidase activity (Miller, Experiments in Molecular Genetics, CSH (1972), pp. 325-355, the disclosure of which is hereby incorporated by reference) were measured for each fraction. The eluting profiles are illustrated in Figure 5. Figure 6 shows that more than 95% of OD280 absorbing material in the crude extracts passed through the column. Only less then 1% was retained by the column and could be eluted with 10mM maltose buffer. In contrast more than 70% of B-galactosidase activity was retained by the column and eluted with 10mM maltose (Figs. 5 and 6). When the pass through fractions were pooled and reapplied onto another cross-linked amylose column, the B-galactosidase activity present in these fractions was not retained. This suggests that a small portion of the hybrid polypeptide was degraded to such a degree that the degraded products lost binding activity with cross-linked amylose, but still maintained some B-galactosidase, enzymatic activity. When the maltose eluted fractions were dialysed and pooled and reapplied onto another cross-linked amylose column, the B-galactosidase activity present in these fractions was, retained and could be eluted with 10mM maltose buffer.

Polyacrylamide Gel Electrophoresis

Affinity chromatography peaks were pooled separately. The maltose eluted peak was concentrated 25-50 fold. 20-40 ul of concentrated sample were mixed with double strength loading buffer (0.5 M Tris-HCl, pH 6.8, 30% glycerol, 4% SDS, 6% beta-mercaptoethanol, 0.4% bromophenol blue) and boiled for two minutes. Samples were applied onto 7 or 10% polyacrylamide gel (29:1). The electrophoresis buffer system was used as described by Laemmli, Nature, Vol. 227, pp. 680-685 (1970), the disclosure of which is hereby incorporated by reference. The gel electrophoresis was performed at 7-10 V/cm or 20 mA for 5 to 7 hours followed by staining with Coomasie Brillant blue R 250 (0.1% coomasie blue, 50% methanol, 10% acetic acid. The gels were destained with destaining solution of 10% acetic acid and 10% methanol).

The results of SDS gel electrophoresis are shown in Figure 7. It appeared that almost all of the protein in the crude extract passed through the column. Only the hybrid polypeptide and small particles of its degraded products were retained by the column and eluted with maltose buffer. The main band on the gel represents the hybrid polypeptide whose molecular weight is estimated at 156k, corresponding to that deduced from the gene fusion sequence.

Native protein gel analysis was also carried out. For native gels the SDS was omitted from the electrophoresis buffer system and the electrophoresis gel was rinsed with water then covered with Z buffer 0.1M NaP04 pH 7.0, KCl 0.01M, Mg2SO4, 0.001M, B-Mercaptoethanol 0.05M) containg XG 20 ug/ml and incubated for 4 hours at 37°C without shaking. When the blue band appeared on gel, the buffer was discarded. This shows that the hybrid polypeptide, which migrated slower than the native B-galactosidese, represents the B-galactosidase enzymatic activity in the maltose buffer eluted fraction (Figure 8).

Immunodiffusion Experiment

Double immunodiffusion (Ouchterlony) experiment was performed on 1% agarose gel in the buffer 10mM Tris, pH 7.2 150mM NaCl. 5-10 ug of sample protein were used (Anti MBP sera obtained from Jon Beckwith of Harvard Medical School. Anti B-galactosidase sera was obtained from Promega Biotech, WI. The purified hybrid polypeptide formed precipitation lines with both anti MBP sera and anti B-galactosidase sera. Pure B-galactosidase formed a precipitation line only with anti B-galactosidase sera and the maltose binding proteins only with anti MBP sera.

EXAMPLE II

Example II describes the cloning, expression and purification of PstI restriction endonuclease as a product of the Mal E-Pst I restriction gene fusion.

Recombinant DNA

The outline of construction of plasmid pCG410 is illustrated in Fig. 9 and 10.

According to the published DNA sequence of Pst I restriction and modification system described in Walder et al., J. Biol. Chem Vol. 259 No. 12, pp. 8015-8026 (1984), the disclosure of which is hereby incorporated by reference, the restriction gene and the methylase gene are transcribed divergently from the promoter region be-

tween the two genes. There is a Hinc II restriction enzyme cleavage site at the eighth codon of the Pst I restriction gene. A Hind III DNA fragment (4.0kb) containing Pst I restriction and modification genes has been cloned in the Hind III site of plasmid pBR322. This plasmid is pGW4400.

30 ug of plasmid pGW440 DNA were cleaved with 30 units of Hind III restriction enzyme and 30 units of Pvu II restriction enzyme in 200 ul of Hind III digestion buffer followed by phenol/chloroform extraction and ethanol precipitation. The DNA was dissolved in 50 ul of TE buffer followed by mixing with 10 ul of loading buffer. A sample was elecrophoresed through 1% of low gelling temperature agarose. After electrophoresis the gel was stained with ethidium bromide and the DNA bands were visualized with UV irradiation as described in Example I. Three bands appeared on gel. The topmost one (4.0kb) was cut out and the DNA was extracted from gel as described in Example I. The purified DNA fragment was ligated with 50 units of T4 DNA Ligase in 0.5 ml of ligation buffer followed by phenol/chloroform extraction and ethanol precipitation. The DNA was cleaved with 30 units of Hinc II restriction enzyme in 100 ul of Hinc digestion buffer followed by phenol/chloroform extraction and ethanol precipitation. The DNA was dissolved in 20 ul of DNA buffer.

5 ug of plasmid pUC18 DNA was cleaved with 10 units of Hinc II restriction enzyme followed by phenol/chloroform extraction and ethanol precipitation. The DNA was dissolved in 10 ul of DNA buffer.

Two DNA preparations, the 4.0 kb fragment from pGW4400 and the Hinc II cleaved pUC-18, were pooled, followed by adding 5 ul of 6x ligation buffer and 2 ul (or 10 units) of T4 ligase and incubated overnight at room temperature. The ligation solution was used to transform competent cells of JM 101 as described in Example I. The transformation mixture was plated on LB plates containing ampicillin 100 ug/ml, XG 20 ug/ml and IPTG 10-4M. After overnight incubation about 100 transformants were obtained. 20% of them were white. 32 white transformants were purified and DNA minipreparations were made from the white transformants as described in Example I. The recombinant plasmids were identified by restriction enzyme analysis. One recombinant plasmid was chosen as pCG228 whose construction is presented in Figure 9.

10-20 ug of plasmid pCG228 DNA purified with BND cellulose were cleaved with 20 units of BamH I restriction enzyme and 20 units of Hind III restriction enzyme in 100 ul of the BamH I-Hind III double digestion buffer (10mM Nacl, 3mM dithiothrietol 10mM MgCl2). The 1.6 kb BamHi-HindIII DNA fragment contained the Pst I restriction gene whose promoter and first 7 codons had been replaced by a BamHi-XbaI-SalI polylinker. This fragment was purified from low gelling temperature agarose gel as described in Example I. The purified DNA fragment was dissolved in 10 ul of DNA buffer.

10 ug of plasmid pCG150 were cleaved with BamH I and Hind III restriction enzymes followed by phenol/chloroform extraction and ethanol precipitation as described above. The DNA was dissolved in 10 ul of DNA buffer.

The two DNA preparations, the 1.6 kb BamH I-Hind III fragment and pCG150 cleaved vector, were pooled and ligated with 10 units of T4 DNA Ligase in 30 ul of ligation buffer by incubation of the ligation solution overnight at 16°C. The ligation solution was used to transform competent cells of MC4100 habouring plasmid pACYC184 (Lac I),. pACYC184 (Lac I) (Chang, et al., J. Bact. Vol.134 No.3 pp.1141-1156 (1978), the disclosure of which is hereby incorporated by reference) is a multicopy plasmid and is compatible with plasmid pBR322 in E. coli K12. A DNA fragment containing the Lac I gene was inserted into the EcoR I cutting site of pACYC184. This is plasmid pACYC184 (Lac I). In order to prepare competent cells of MC4100 harbouring pACYC184 (Lac I), MC4100 was first transformed with plasmid pACYC184 (Lac I). The transformants (tetracycline resistant) were then used to prepare competent cells as described in Example I. These are competent cells of MC4100 harbouring pACYC184 (Lac I). The transformation mixture was placed onto LB plates containing ampicillin, 100 ug/ml, tetracycline 20 ug/ml. About 50-100 transformants appeared on each plate after overnight incubation. The plates were replicated onto LB plates containing ampicillin 100 ug/ml, tetracycline 20 ug/ml and IPTG 4x10-4 M. The replicated plates were incubated overnight at 37°C. The transformants which grew on LB-ampicillin-tetracycline plates but failed to grow on LB-ampicillin-tetracycline-IPTG plates were saved and purified on LB-ampicillin-tetracycline plates. DNA mini-preparations were made from the IPTG sensitive transformants and used to transform JM103 or 71-18. The transformants which were resistent to ampicillin but sensitive to tetracycline and 10^{-5}M IPTG were saved. DNA mini preparations were made from these IPTG sensitive transformants and analyzed with restriction enzyme digestions. One recombinant plasmid was chosen as pCG410 whose construction is presented in Figure 10.

Affinity Chromatography of Pst I - Mal E Fusion

E. coli strain MC4100 harbouring both plasmids pCG410 and pACYC184 (Lac I) was cultivated to late log phas in rich medim containing ampicillin 100 ug/ml and tetracycline 20 ug/ml at 37°C. IPTG was added to 4 x 10-4 M and the culture was incubated for additonal 1.5 hours at 37°C. The cells were harvested and the cellular crude extract was prepared as described in Example I. The cellular extract was applied to a cross-linked amylose column and affinity chromatography was performed as described in Example I. More than 99% of (OD 280) absorbing material in the cellular crude extract passed through cross-linked amylose column. Less than 1% of OD 280 absorbing material bound to the column could

be eluted with the maltose buffer. Pst I restriction enzymatic activity was found in the pass through fraction and in the maltose buffer eluted fractions. High levels of non-specific DNAase were found in the pass through fraction but not in the maltose buffer eluted fractions. The pass through fractions consisting of the main protein peak were pooled and applied onto another cross-linked amylose column. Neither protein nor DNAase acitivity, including Pst I restriction like activity, were found to be retained by the column. In contrast, when the Pst I restriction like enzymatic activity in the maltose eluted fractions was pooled, dialysed and reapplied onto another cross-linked amylose column, all of the activity was retained by column and could be eluted with maltose buffer.

Polyacrylamide Gel Electrophoresis

The fractions consisting of the main protein peak and the maltose eluted peak were pooled seperately. The maltose eluted pool was concentrated 25-50 fold as described in Example I. The pooled samples above were used for SDS polyacrylamide gel electrophoresis as described in Example I. The results are shown in Figure 11. Three proteins were eluted with the maltose buffer as determined by the SDS gel. The topmost band represents a protein whose molecular weight is estimated at 78 K daltons corresponding to that deduced from the sequence of the MalE-PstI gene-fusion. The lowest band comigrated with native maltose binding protein and was believed to represent the product of the Mal E gene of the host cell. It is also possible that this represents the degraded product from the hybrid polypeptide, formed as a protease resistant domain in the hybrid polypeptide. The third band which migrated slightly slower than either MBP or Pst I proteins may be degradation products.

Example III

Preparation of Immobilized Protein Bioreactor.

Ten milliliters of late log phase culture of strain SF1362 harboring plasmid pCG325 was harvested by centrifugation. The cell pellet was suspended in 2 ml. of buffer (10mM TriS-HCl pH 7.2). Crude extract was prepared as described in Example I. The cell extract was applied to a 0.6 x 2.5 cm cross-linked amylose column, and washed with buffer as in Example I.

Cleavage of ONPG by the Bioreactor.

The bioreactor column was equilibrated with Z buffer as in Example I at room temperature. 500 ml of Z buffer containing 0.1% ONPG was applied to the column at room temperature with a flow rate of 0.5 ml/min. The pass through fraction was collected and the conversion to ONPG to ONP and free sugar was determined to be greater than 95%. After use the bioreactor may washed with Z buffer and stored at 4 degrees centigrade. The bioreactor can be reused multiple times.

Claims

1. A method for producing and purifying a target protein molecule comprising:

   (a) constructing a DNA expression vector which expresses a hybrid polypeptide in a transformed host cell, the hybrid polypeptide comprising the target protein molecule and a non-enzymatic biologically functional sugar binding protein, having a specific affinity for a substrate which binds to the non-enzymatic biologically functional sugar binding protein; and

   (b) introducing the expression vector into an appropriate host cell and expressing the hybrid polypeptide;

   (c) contacting the hybrid polypeptide produced by the transformed cell with the substrate to which the non-enzymatic biologically functional sugar binding protein binds; and

   (d) recovering the target protein molecule.

2. The method of claim 1, wherein the DNA encoding the hybrid polypeptide contains a linking DNA fragment which links the DNA encoding the protein molecule with the DNA encoding the non-enzymatic biologically functional sugar binding protein.

3. The method of claim 1, wherein the non-enzymatic biologically functional sugar binding protein is maltose binding protein and the substrate is selected from the group consisting of maltose, maltodextrins and macromolecular alpha (1→4) linked glucans.

4. The method of claim 3, wherein the substrate is crosslinked amylose.

5. The method of claim 1, comprising the further step of releasing the hybrid polypeptide from the substrate by contacting the bound hybrid polypeptide with a substance which displaces the hybrid polypeptide.

6. The method of claim 1, wherein the substrate is contained within an affinity column.

7. The method of claim 1, comprising the further step of cleaving the protein molecule from the hybrid polypeptide.

8. The method of claim 1 or 2 wherein :

the hybrid polypeptide comprises the protein molecule, a maltose binding protein or portion thereof having a specific affinity for a substrate which binds to the maltose binding protein, and a linking sequence interposed between said protein molecule and said maltose binding protein or portion thereof, said linking sequence having a Factor Xa protease cleavage site.

9. A fusion vector which comprises:

(a) a DNA fragment coding for a non-enzymatic biologically functional sugar binding protein, the non-enzymatic biologically functional sugar binding protein having a specific affinity for a substrate which binds to the non-enzymatic biologically functional sugar binding protein; and

(b) a DNA fragment which codes for a linking sequence for linking the DNA coding for the non-enzymatic biologically functional sugar binding protein or portion thereof with a target protein molecule.

10. The fusion vector of claim 9, wherein the non-enzymatic biologically functional sugar binding protein is maltose binding protein and the substrate is selected from the group consisting of maltose, maltodextrins and macromolecular alpha (1→4) linked glucans.

11. The fusion vector of claim 9, wherein the linking sequence comprises one or more restriction sites.

12. The fusion vector of claim 9, wherein the linking sequence codes for a polypeptide which is recognized and cleaved by a proteolytic agent.

13. The fusion vector of claim 9, wherein the linking sequence codes for a spacer polypeptide which separates the non-enzymatic biologically functional sugar binding protein from the target protein molecule.

14. The fusion vector of claim 9, comprising the plasmid pCG150 obtainable from the American Type Culture Collection Deposit No. 67345.

15. A fusion vector according to claim 9 for constructing an expression vector which expresses a maltose binding protein fused to a protein molecule to be purified, comprising:

(a) a DNA fragment coding for the maltose binding protein or biologically active portion thereof, the maltose binding protein having a specific affinity for a substrate which binds to the maltose binding protein; and

(b) a DNA fragment which codes for a linking sequence having a Factor Xa protease cleavage site, wherein said DNA fragment is adapted for linking the DNA coding the maltose binding protein with the DNA coding for the protein molecule.

16. A DNA expression vector for producing a purified target protein molecule, which upon expression produces a non-enzymatic biologically functional sugar binding protein fused to the target protein molecule comprising:

(a) a DNA fragment coding for the non-enzymatic biologically functional sugar binding protein, the sugar binding protein having a specific affinity for a substrate which binds to the sugar binding protein; and

(b) a DNA fragment coding for the target protein molecule.

17. The expression vector of claim 16, wherein the non-enzymatic biologically functional sugar binding protein is maltose binding protein and the substrate is selected from the group consisting of maltose, maltodextrins and macromolecular (1→4) linked glucans.

18. The expression vector of claim 16, wherein a DNA fragment coding for a linking sequence is interposed between the DNA encoding the non-enzymatic biologically functional sugar binding protein and the DNA encoding the protein molecule.

19. The expression vector of claim 18, wherein the linking sequence comprises one or more restriction sites.

20. The expression vector of claim 18, wherein the linking sequence codes for a polypeptide which is recognized and cleaved by a proteolytic agent.

21. The expression vector of claim 18, wherein the linking sequence codes for a spacer polypeptide which separates the binding protein from the protein molecule expressed by the expression vector.

22. The DNA expression vector of claim 16 or 18, which upon expression produces a maltose binding protein fused to the protein molecule, comprising:

a DNA fragment coding for the maltose binding protein or biologically active portion thereof, the maltose binding protein having a specific affinity for a substrate which binds to the maltose binding protein; and

a linking DNA fragment coding for a linking sequence interposed between said first and second DNA fragments, wherein said linking sequence contains a Factor Xa protease cleavage site.

Patentansprüche

1. Verfahren zum Herstellen und Reinigen eines Zielproteinmoleküls, umfassend:

   (a) Konstruieren eines DNA-Expressionsvektors, welcher ein Hybridpolypeptid in einer transformierten Wirtszelle exprimiert, wobei das Hybridpolypeptid das Zielproteinmolekül und ein nicht-enzymatisches biologisch funktionelles zuckerbindendes Protein umfaßt, das eine spezifische Affinität für ein Substrat aufweist, welches an das nicht-enzymatische biologisch funktionelle zuckerbindende Protein bindet; und

   (b) Einführen des Expressionsvektors in eine geeignete Wirtszelle und Exprimieren des Hybridpolypeptids;

   (c) In-Kontakt-Bringen des Hybridpolypeptids, welches von der transformierten Zelle produziert wurde, mit dem Substrat, an welches das nicht-enzymatische biologisch funktionelle zukkerbindende Protein bindet; und

   (d) Gewinnen des Zielproteinmoleküls.

2. Verfahren nach Anspruch 1, wobei die DNA, welche das Hybridpolypeptid codiert, ein Verknüpfungs-DNA-Fragment enthält, welches die DNA, die das Proteinmolekül codiert, mit der DNA, die das nicht-enzymatische biologisch funktionelle zuckerbindende Protein codiert, verknüpft.

3. Verfahren nach Anspruch 1, wobei das nicht-enzymatische biologisch funktionelle zuckerbindende Protein Maltosebindendes Protein ist und das Substrat ausgewählt wird aus der Gruppe bestehend aus: Maltose, Maltodextrinen und makromolekularen alpha(1→4)-verknüpften Glucanen.

4. Verfahren nach Anspruch 3, wobei das Substrat quervernetzte Amylose ist.

5. Verfahren nach Anspruch 1, welches den weiteren Schritt des Freisetzens des Hybridpolypeptids von dem Substrat durch In-Kontakt-Bringen des gebundenen Hybridpolypeptids mit einer Substanz, welche das Hybridpolypeptid verdrängt, umfaßt.

6. Verfahren nach Anspruch 1, wobei das Substrat innerhalb einer Affinitätssäule enthalten ist.

7. Verfahren nach Anspruch 1, welches den weiteren Schritt des Abspaltens des Proteinmoleküls von dem Hybridpolypeptid umfaßt.

8. Verfahren nach Anspruch 1 oder 2, wobei das Hybridpolypeptid das Proteinmolekül, ein Maltose-bindendes Protein oder einen Teil davon mit einer spezifischen Affinität für ein Substrat, welches an das Maltose-bindende Protein bindet, und eine Verknüpfungssequenz umfaßt, die zwischen dem Proteinmolekül und dem Maltose-bindenden Protein oder einem Teil davon, eingefügt ist, wobei die Verknüpfungssequenz eine Faktor Xa-Protease-Spaltungsstelle aufweist.

9. Fusionsvektor, welcher umfaßt:

   (a) ein DNA-Fragment, welches für ein nicht-enzymatisches biologisch funktionelles zuckerbindendes Protein codiert, wobei das nicht-enzymatische biologisch funktionelle zukkerbindende Protein eine spezifische Affinität zu einem Substrat aufweist, welches an das nicht-enzymatische biologisch funktionelle zukkerbindende Protein bindet; und

   (b) ein DNA-Fragment, welches für eine Verknüpfungssequenz codiert, um die DNA, welche für das nicht-enzymatische biologisch funktionelle zuckerbindende Protein oder einen Teil davon codiert, mit einem Zielproteinmolekül zu verknüpfen.

10. Fusionsvektor nach Anspruch 9, wobei das nicht-enzymatische biologisch funktionelle zuckerbindende Protein Maltose-bindendes Protein ist und das Substrat ausgewählt wird aus der Gruppe bestehend aus: Maltose, Maltodextrinen und makromolekularen alpha(1-4)-verknüpften Glucanen.

11. Fusionsvektor nach Anspruch 9, wobei die Verknüpfungssequenz eine oder mehrere Restriktionsstellen umfaßt.

12. Fusionsvektor nach Anspruch 9, wobei die Verknüpfungssequenz für ein Polypeptid codiert, welches durch ein proteolytisches Mittel erkannt und gespalten wird.

13. Fusionsvektor nach Anspruch 9, wobei die Verknüpfungssequenz für ein Spacer-Polypeptid codiert, welches das nicht-enzymatische biologisch funktionelle zuckerbindende Protein von dem Zielproteinmolekül trennt.

14. Fusionsvektor nach Anspruch 9, welcher das Plas-

mid pCG150 umfaßt, welches von der American Type Culture Collection unter der Hinterlegungsnr. 67345 erhältlich ist.

15. Fusionsvektor nach Anspruch 9 zum Konstruieren eines Expressionsvektors, welcher ein Maltose-bindendes Protein exprimiert, das mit einem zu reinigenden Proteinmolekül verbunden ist, umfassend:

(a) ein DNA-Fragment, welches für das Maltose-bindende Protein oder einen biologisch aktiven Teil davon codiert, wobei das Maltose-bindende Protein eine spezifische Affinität für ein Substrat aufweist, das an das Maltose-bindende Protein bindet; und

(b) ein DNA-Fragment, welches für eine Verknüpfungssequenz mit einer Faktor Xa-Protease-Spaltungsstelle codiert, wobei das DNA-Fragment angepaßt ist, um die DNA, die das Maltose-bindende Protein codiert, mit der DNA, die für das Proteinmolekül codiert, zu verknüpfen.

16. DNA-Expressionsvektor zum Herstellen eines gereinigten Zielproteinmoleküls, welches unter Expression ein nicht-enzymatisches biologisch funktionelles zuckerbindendes Protein produziert, das mit dem Zielproteinmolekül verbunden ist, umfassend:

(a) ein DNA-Fragment, welches für das nicht-enzymatische biologisch funktionelle zukkerbindende Protein codiert, wobei das zuckerbindende Protein eine spezifische Affinität zu einem Substrat aufweist, das an das zuckerbindende Protein bindet; und

(b) ein DNA-Fragment, welches für das Zielproteinmolekül codiert.

17. Expressionsvektor nach Anspruch 16, wobei das nicht-enzymatische biologisch funktionelle zuckerbindende Protein Maltose-bindendes Protein ist und das Substrat ausgewählt wird aus der Gruppe bestehend aus: Maltose, Maltodextrinen und makromolekularen alpha(1→4)-verknüpften Glucanen.

18. Expressionsvektor nach Anspruch 16, wobei ein DNA-Fragment, das für eine Verknüpfungssequenz codiert, zwischen der DNA, die das nicht-enzymatische biologisch funktionelle zuckerbindende Protein codiert, und der DNA, die das Proteinmolekül codiert, eingefügt ist.

19. Expressionsvektor nach Anspruch 18, wobei die Verknüpfungssequenz eine oder mehrere Restriktionsstellen umfaßt.

20. Expressionsvektor nach Anspruch 18, wobei die Verknüpfungssequenz für ein Polypeptid codiert, das durch ein proteolytisches Mittel erkannt und gespalten wird.

21. Expressionsvektor nach Anspruch 18, wobei die Verknüpfungssequenz für ein Spacer-Polypeptid codiert, welches das bindende Protein von dem Proteinmolekül, das durch den Expressionsvektor exprimiert wird, trennt.

22. DNA-Expressionsvektor nach Anspruch 16 oder 18, welcher unter Expression ein Maltose-bindendes Protein produziert, das mit dem Proteinmolekül verbunden ist, umfassend:

ein DNA-Fragment, welches für das Maltose-bindende Protein oder einen biologisch aktiven Teil davon codiert, wobei das Maltose-bindende Protein eine spezifische Affinität für ein Substrat aufweist, das an das Maltose-bindende Protein bindet; und

ein Verknüpfungs-DNA-Fragment, welches für eine Verknüpfungssequenz codiert, die zwischen dem ersten und zweiten DNA-Fragment eingefügt ist, wobei die Verknüpfungssequenz eine Faktor Xa-Protease-Spaltungsstelle enthält.

Revendications

1. Méthode pour la production et la purification d'une molécule de protéine cible comprenant les étapes suivantes :

(a) la construction d'un vecteur d'expression d'ADN qui exprime un polypeptide hybride dans une cellule hôte transformée, le polypeptide hybride comprenant la molécule de protéine cible, et une protéine non enzymatique fixatrice de sucre biologiquement fonctionnelle, ayant une affinité spécifique pour un substrat qui se fixe à une protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre ; et
(b) l'introduction du vecteur d'expression dans un hôte cellulaire approprié et l'expression du polypeptide hybride ;
(c) la mise en contact du polypeptide hybride produit par la cellule transformée avec le substrat auquel la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre se fixe ; et
(d) la récupération de la molécule de protéine cible.

2. Méthode selon la revendication 1, où l'ADN codant

pour le polypeptide hybride contient un fragment d'ADN de liaison qui relie l'ADN codant pour la molécule de protéine avec l'ADN codant pour la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre.

3. Méthode selon la revendication 1, où la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre est la protéine fixant le maltose et le substrat est choisi parmi le groupe constitué du maltose, des maltodextrines et de glycanes macromoléculaires liés en alpha (1- 4).

4. Méthode selon la revendication 3, où le substrat est l'amylose réticulée.

5. Méthode selon la revendication 1, comprenant l'étape supplémentaire de relarguage du polypeptide hybride du substrat par la mise en contact du polypeptide hybride fixé avec une substance qui déplace le polypeptide hybride.

6. Méthode selon la revendication 1 où le substrat est contenu dans une colonne d'affinité.

7. Méthode selon la revendication 1, comprenant l'étape supplémentaire de clivage de la molécule de protéine du polypeptide hybride.

8. Méthode selon la revendication 1 ou 2, où le polypeptide hybride comprend une molécule de protéine, une protéine fixant le maltose ou un fragment de celle-ci ayant une affinité spécifique pour un substrat qui se fixe à la protéine fixant le maltose, et une séquence de liaison interposée entre ladite molécule de protéine et ladite protéine fixant le maltose ou son fragment, ladite séquence de liaison ayant un site de clivage de la protéase Facteur Xa.

9. Vecteur de fusion comprenant :

(a) un fragment d'ADN codant pour une protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre, la protéine non enzymatique biologiquement fonctionnelle fixatrice du sucre ayant une affinité spécifique pour un substrat qui se fixe à la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre ; et
(b) un fragment d'ADN qui code pour une séquence de liaison pour relier l'ADN codant pour la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre ou une portion de celle-ci avec une molécule de protéine cible.

10. Vecteur de fusion selon la revendication 9, où la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre est la protéine fixant le maltose et le substrat est choisi parmi le groupe constitué du maltose, des maltodextrines et des glycanes macromoléculaires liés en alpha (1-4).

11. Vecteur de fusion selon la revendication 9, où la séquence de liaison comprend un ou plusieurs sites de restriction.

12. Vecteur de fusion selon la revendication 9, où la séquence de liaison code pour un polypeptide qui est reconnu et clivé par un agent protéolytique.

13. Vecteur de fusion selon la revendication 9, où la séquence de liaison code pour un polypeptide espaceur qui sépare la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre de la molécule de protéine cible.

14. Vecteur de fusion selon la revendication 9, comprenant le plamisde pCG150 déposé à l'American Type Culture Collection sous le numéro 67345.

15. Vecteur de fusion selon la revendication 9, pour la construction d'un vecteur d'expression qui exprime une protéine fixatrice de maltose fusionnée à une molécule de protéine devant être purifiée, comprenant :

(a) un fragment d'ADN codant pour la protéine fixatrice de maltose ou une portion de celle-ci biologiquement active, la protéine fixatrice de maltose ayant une affinité spécifique pour un substrat qui se fixe à la protéine fixatrice du maltose ; et
(b) un fragment d'ADN qui code pour une séquence de liaison ayant un site de clivage pour la protéase Facteur Xa où ledit fragment d'ADN est adapté pour relier l'ADN codant pour la protéine fixatrice de maltose avec l'ADN codant pour la molécule de protéine.

16. Vecteur d'expression d'ADN pour la production d'une molécule de protéine cible purifiée, qui produit lors de son expression une protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre fusionnée à une molécule de protéine cible comprenant :

(a) un fragment d'ADN codant pour une protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre, la protéine fixatrice de sucre ayant une affinité spécifique pour un substrat qui se fixe à la protéine fixatrice de sucre, et
(b) un fragment d'ADN codant pour une molécule de protéine cible.

17. Vecteur d'expression selon la revendication 16 où la protéine non enzymatique biologiquement fonction-

nelle fixatrice de sucre est la protéine fixant le maltose et le substrat est choisi parmi le groupe constitué du maltose, des maltodextrines et de glycanes macromoléculaires liés en alpha (1-4).

18. Vecteur d'expression selon la revendication 16 où un fragment d'ADN codant pour une séquence de liaison est interposé entre l'ADN codant pour la protéine non enzymatique biologiquement fonctionnelle fixatrice de sucre et l'ADN codant pour la molécule de protéine.

19. Vecteur d'expression selon la revendication 18 où la séquence de liaison comprend un ou plusieurs sites de restriction.

20. Vecteur d'expression selon la revendication 18 où la séquence de liaison code pour un polypeptide qui est reconnu et clivé par un agent protéolytique.

21. Vecteur d'expression selon la revendication 18 où la séquence de liaison code pour un polypeptide espaceur qui sépare la protéine fixatrice de la molécule de protéine exprimée par le vecteur d'expression.

22. Vecteur d'expression d'ADN selon la revendication 16 ou 18 qui, lors de son expression, produit une protéine fixatrice de maltose fusionnée à la molécule comprenant :

un fragment d'ADN codant pour la protéine fixatrice de maltose ou une portion de celle-ci biologiquement fonctionnelle, la protéine fixant le maltose ayant une affinité spécifique pour un substrat qui se fixe à la protéine fixant le maltose ; et
un fragment d'ADN de liaison codant pour une séquence de liaison interposée entre lesdits premier et second fragments d'ADN ou ladite séquence de liaison contient un site de clivage pour la protéase facteur XA.

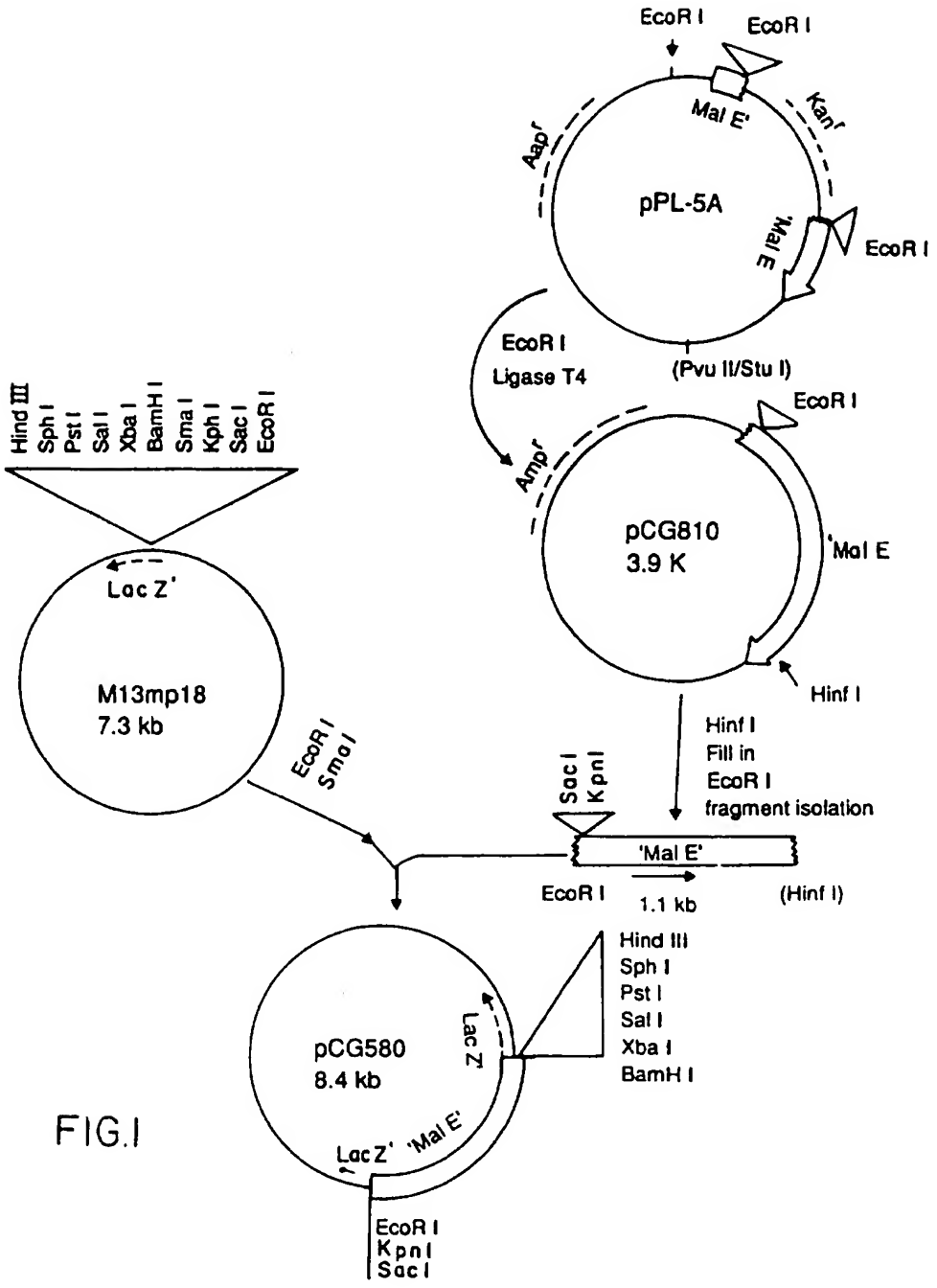

EcoR I
EcoR I
Mal E'
Kan'
Aap'
pPL-5A
'Mal E
EcoR I
(Pvu II/Stu I)
EcoR I
Ligase T4
Amp'
EcoR I
pCG810
3.9 K
'Mal E
Hinf I
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Sma I
Kph I
Sac I
EcoR I
Lac Z'
M13mp18
7.3 kb
EcoR I
Sma I
Hinf I
Fill in
EcoR I
fragment isolation
Sac I
Kpn I
'Mal E'
EcoR I
1.1 kb
(Hinf I)
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Lac Z'
pCG580
8.4 kb
Lac Z'
'Mal E'
EcoR I
Kpn I
Sac I

FIG.1

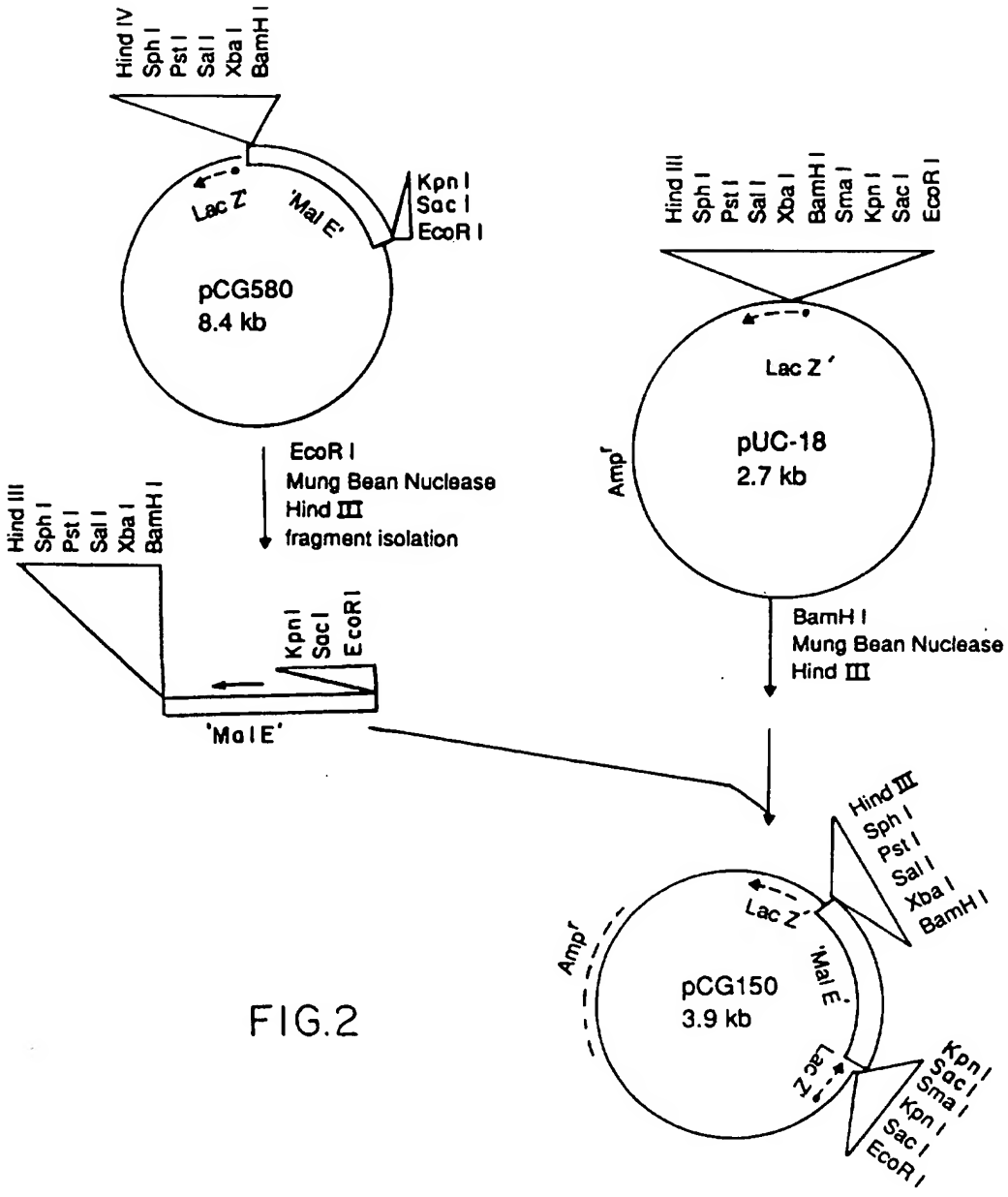

Hind IV
Sph I
Pst I
Sal I
Xba I
BamH I
Lac Z'
'Mal E'
Kpn I
Sac I
EcoR I
pCG580
8.4 kb
EcoR I
Mung Bean Nuclease
Hind III
fragment isolation
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Kpn I
Sac I
EcoRI
'Mal E'
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Sma I
Kpn I
Sac I
EcoR I
Lac Z'
Amp'
pUC-18
2.7 kb
BamH I
Mung Bean Nuclease
Hind III
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Lac Z'
Amp'
'Mal E'
pCG150
3.9 kb
Lac Z'
Kpn I
Sac I
Sma I
Kpn I
Sac I
EcoR I

FIG.2

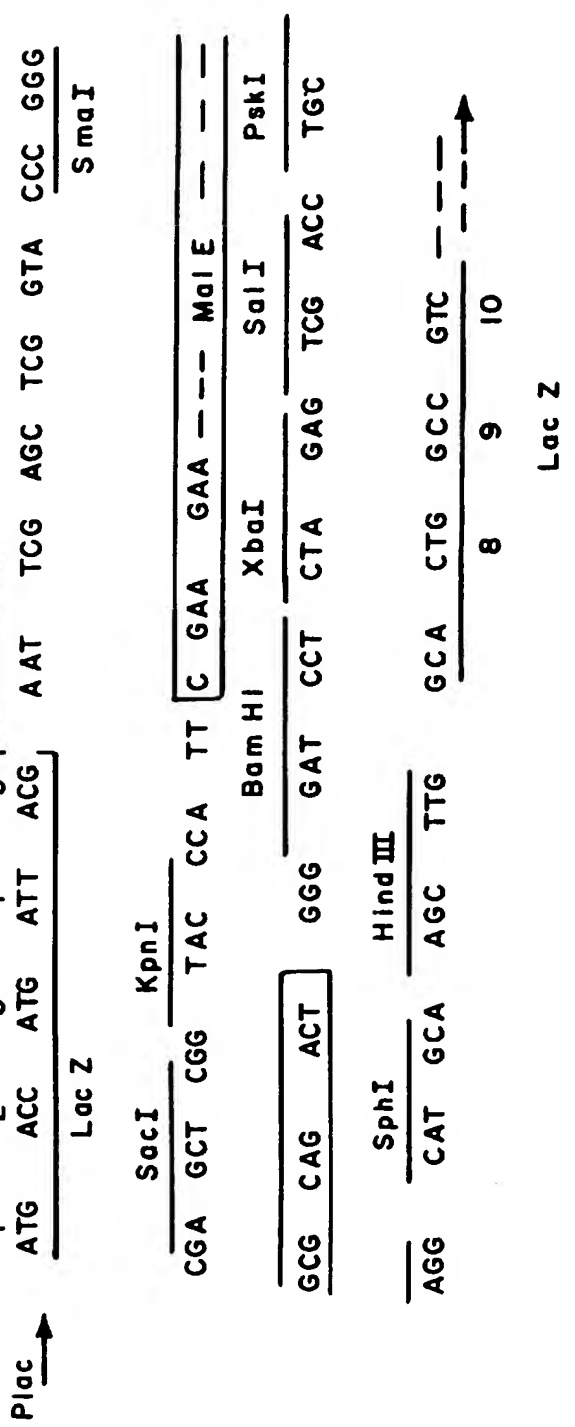

Plac
1 2 3 4 5
ATG ACC ATG ATT ACG
Lac Z
EcoRI
AAT
SacI
TCG AGC TCG
KpnI
GTA CCC GGG
SmaI
SacI
CGA GCT CGG
KpnI
TAC CCA TT
C GAA GAA --- MalE
BamHI
GGG GAT CCT
XbaI
CTA GAG
SalI
TCG ACC
PskI
TGC
GCG CAG ACT
SphI
CAT GCA
HindIII
AGC TTG
GCA CTG GCC GTC ---
8 9 10
Lac Z

FIG.3

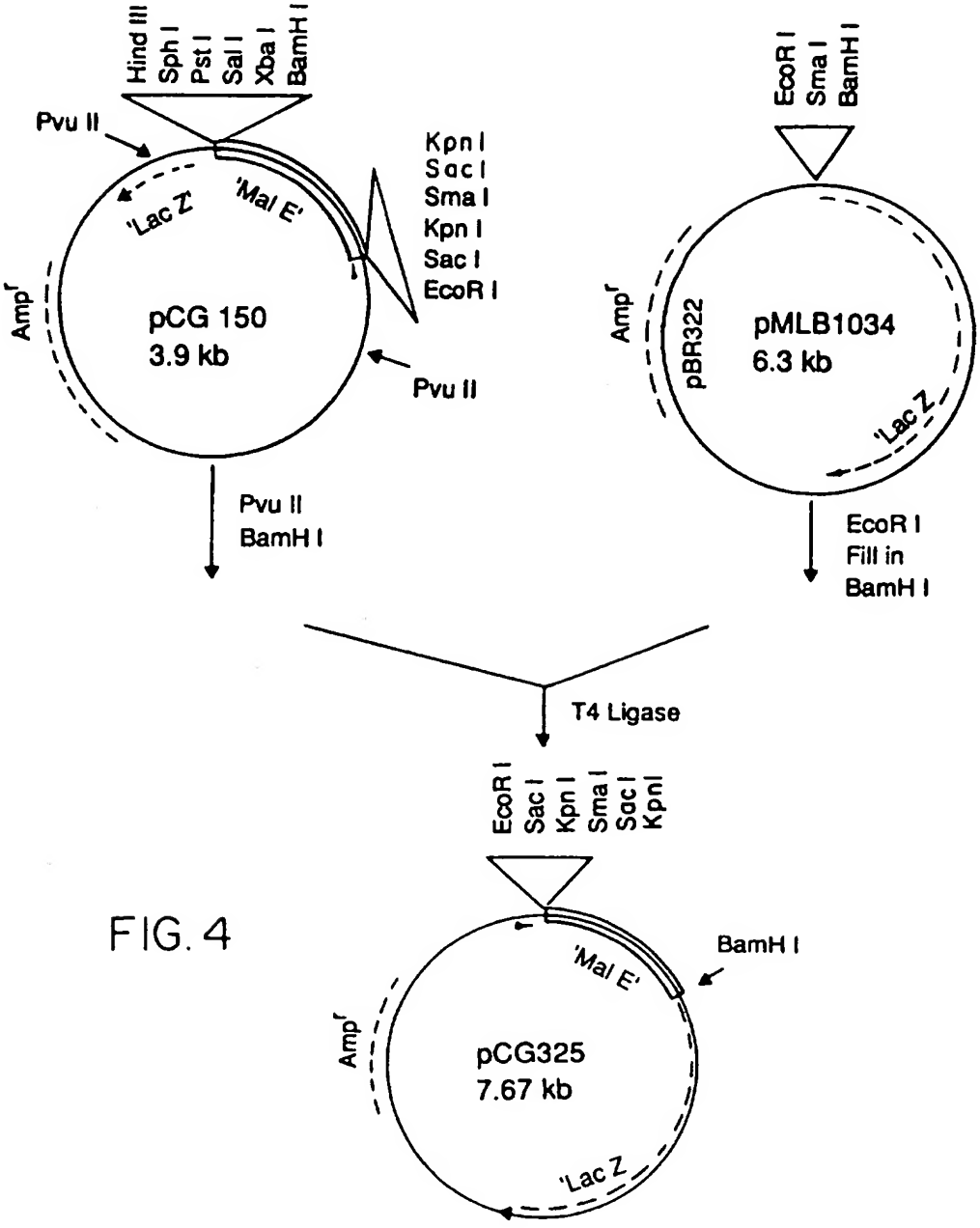

Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Pvu II
'Lac Z
'Mal E'
Amp^r
pCG 150
3.9 kb
Kpn I
Sac I
Sma I
Kpn I
Sac I
EcoR I
Pvu II
Pvu II
BamH I
EcoR I
Sma I
BamH I
Amp^r
pBR322
pMLB1034
6.3 kb
'Lac Z
EcoR I
Fill in
BamH I
T4 Ligase
EcoR I
Sac I
Kpn I
Sma I
Sac I
Kpn I
BamH I
'Mal E'
Amp^r
pCG325
7.67 kb
'Lac Z

FIG. 4

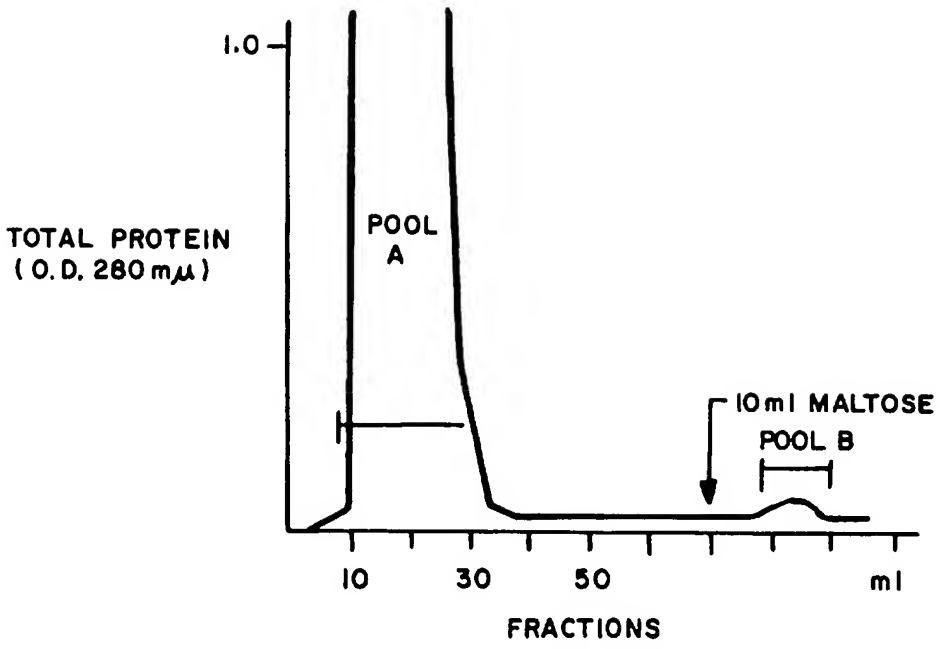

1.0
TOTAL PROTEIN
(O.D. 280 mμ)
POOL
A
10ml MALTOSE
POOL B
10
30
50
ml
FRACTIONS

FIG.5

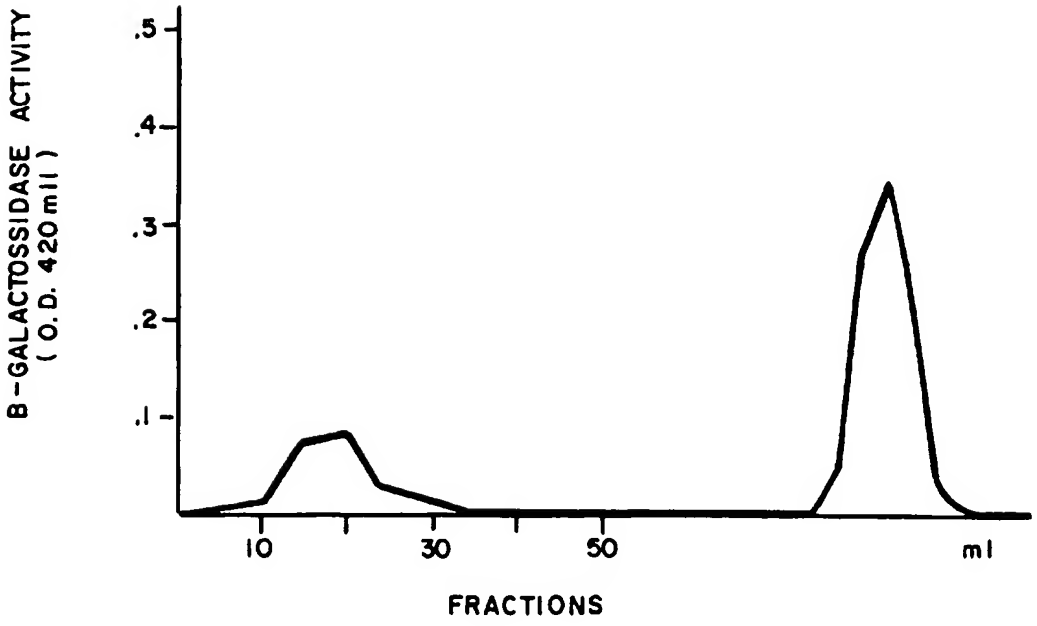

B-GALACTOSSIDASE ACTIVITY
(O.D. 420 mll)
.5
.4
.3
.2
.1
10
30
50
ml
FRACTIONS

FIG.6

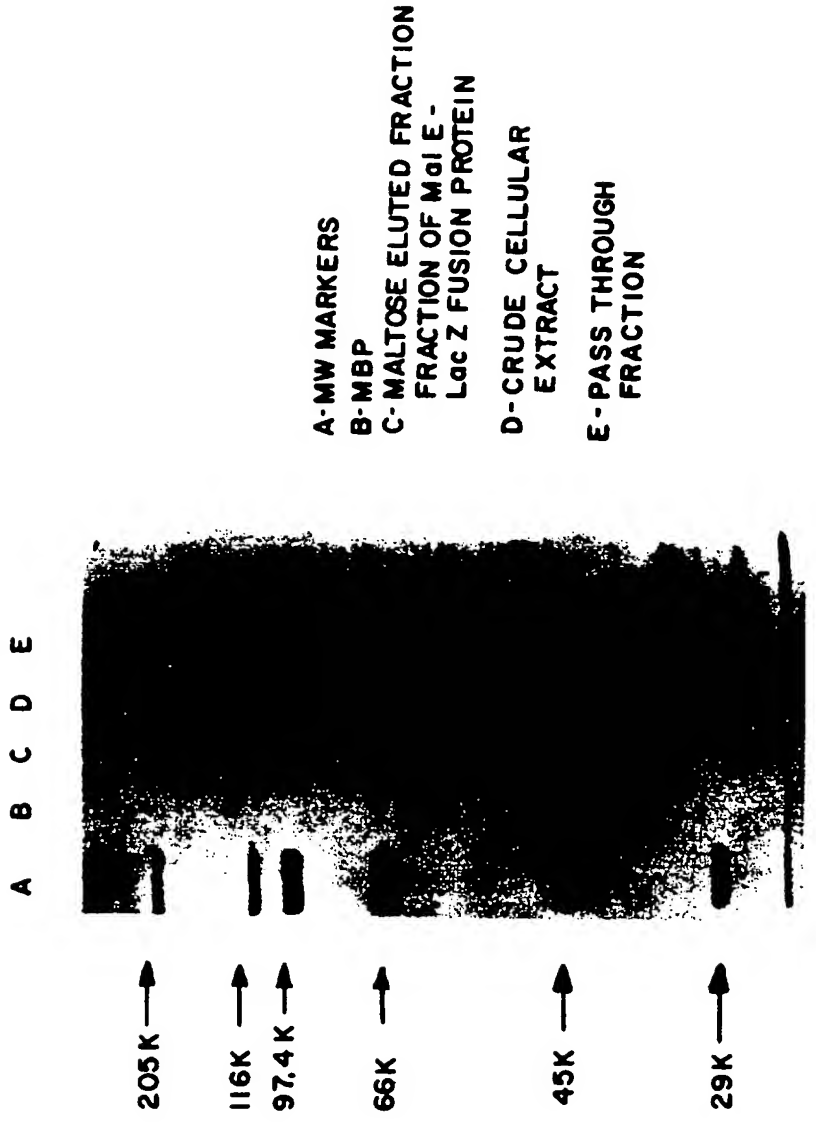
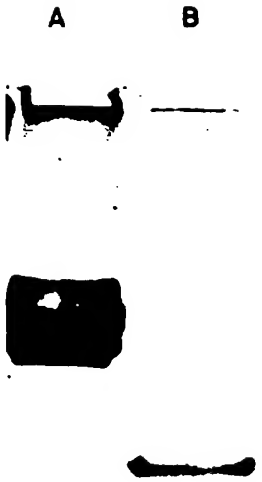

A-MW MARKERS
B-MBP
C-MALTOSE ELUTED FRACTION FRACTION OF Mal E-Lac Z FUSION PROTEIN
D-CRUDE CELLULAR EXTRACT
E-PASS THROUGH FRACTION
A B C D E
205K
116K
97.4K
66K
45K
29K

FIG.7



A.- Mal E - LacZ
FUSION PROTEIN

B.- β - GALACTOSIDASE

FIG.8

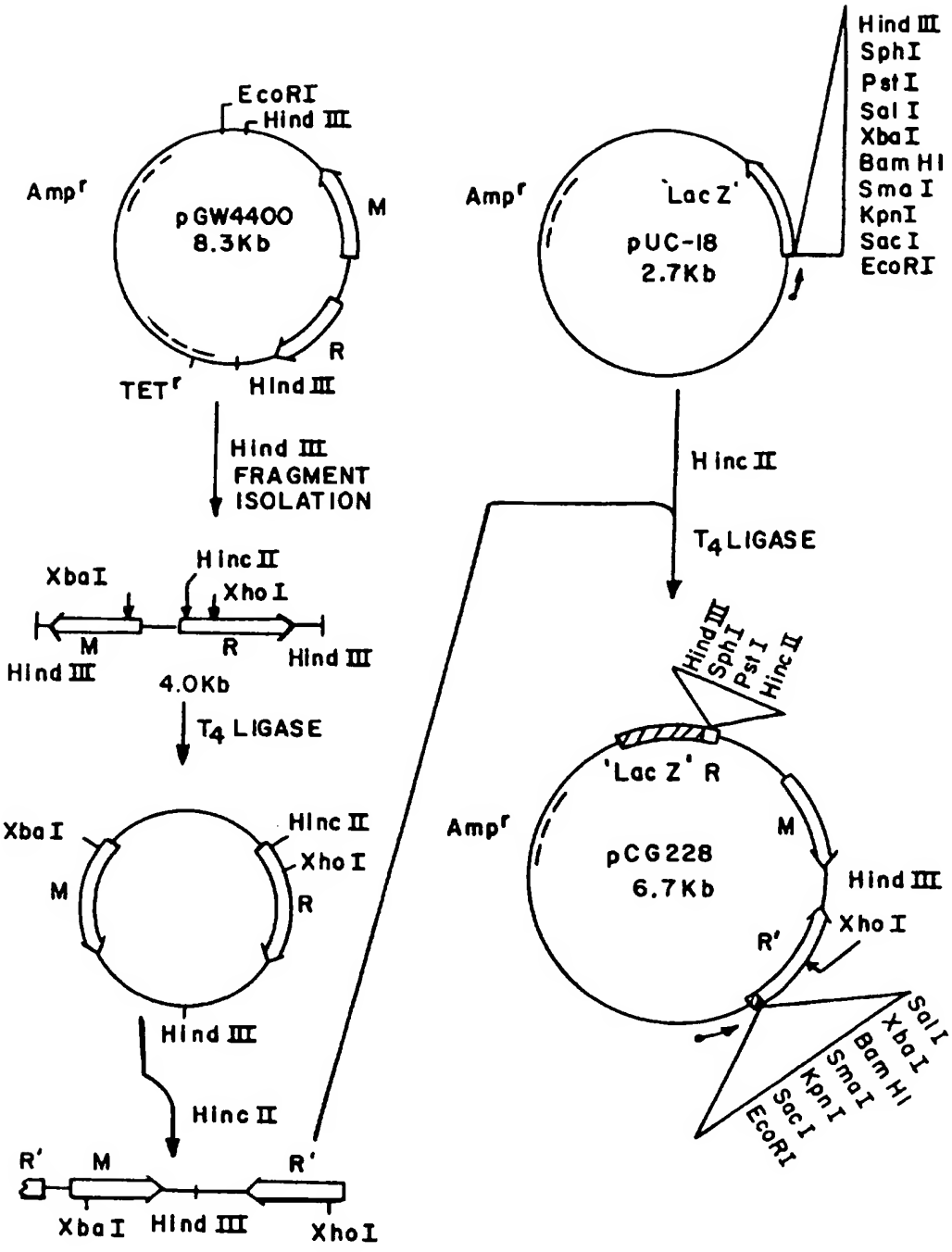

EcoRI
Hind III
Amp^r
pGW4400
8.3Kb
M
R
TET^r
Hind III
Hind III
FRAGMENT
ISOLATION
XbaI
HincII
XhoI
M
R
Hind III
Hind III
4.0Kb
T4 LIGASE
XbaI
HincII
XhoI
M
R
Hind III
HincII
R'
M
R'
XbaI
Hind III
XhoI
Amp^r
'Lac Z'
pUC-18
2.7Kb
Hind III
SphI
PstI
SalI
XbaI
BamHI
SmaI
KpnI
SacI
EcoRI
HincII
T4 LIGASE
Hind III
SphI
PstI
HincII
'Lac Z' R
Amp^r
M
pCG228
6.7Kb
Hind III
R'
XhoI
SalI
XbaI
BamHI
SmaI
KpnI
SacI
EcoRI

FIG. 9

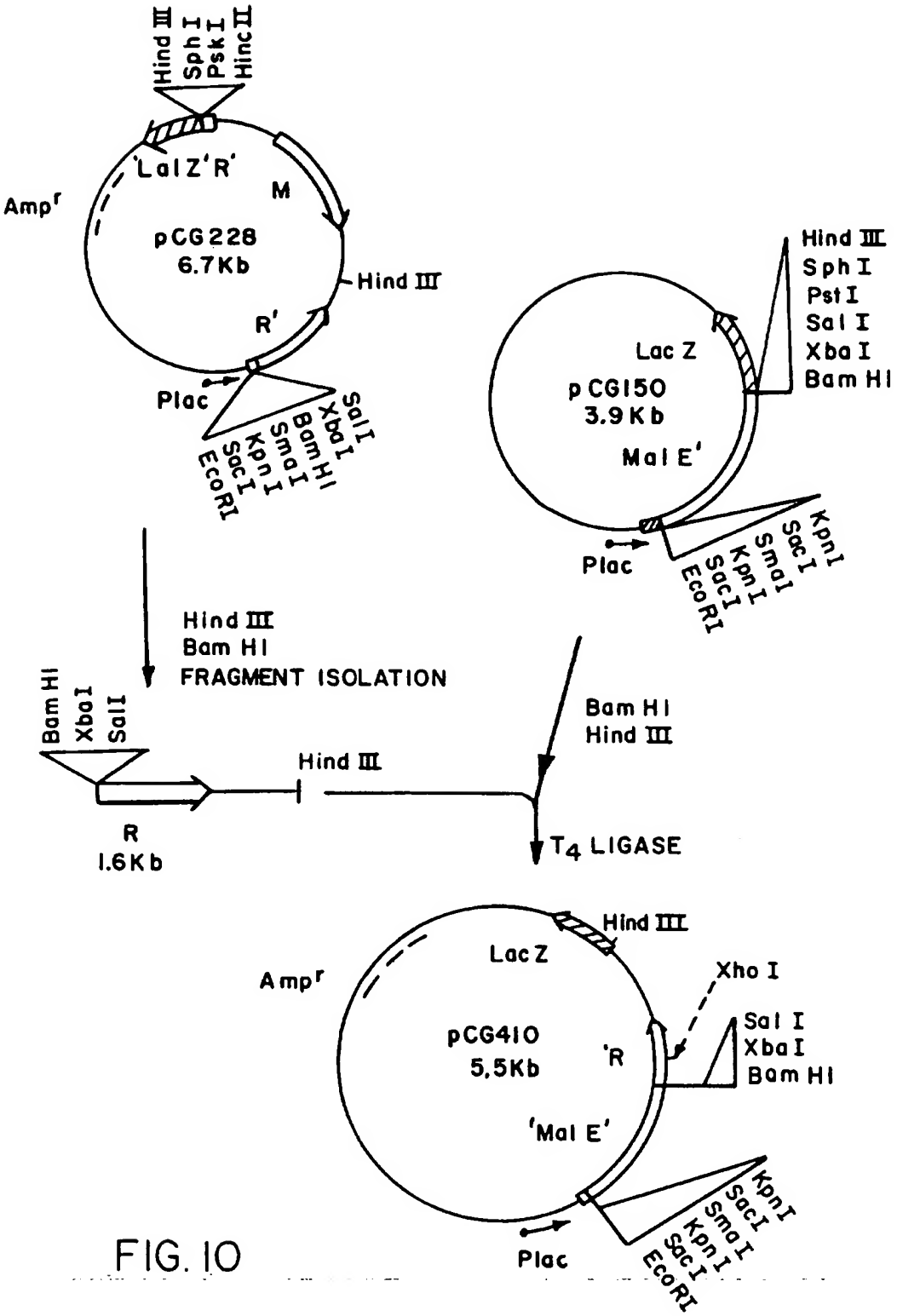

Hind III
Sph I
Pst I
Hinc II
'LacZ'R'
M
Amp^r
pCG228
6.7Kb
Hind III
R'
Plac
Sal I
Xba I
Bam HI
Sma I
Kpn I
Sac I
EcoRI
Lac Z
pCG150
3.9Kb
Mal E'
Hind III
Sph I
Pst I
Sal I
Xba I
Bam HI
Plac
Kpn I
Sac I
Sma I
Kpn I
Sac I
EcoRI
Hind III
Bam HI
FRAGMENT ISOLATION
Bam HI
Xba I
Sal I
Hind III
R
1.6Kb
Bam HI
Hind III
T4 LIGASE
Hind III
Lac Z
Amp^r
Xho I
pCG410
5.5Kb
'R
Sal I
Xba I
Bam HI
'Mal E'
Plac
Kpn I
Sac I
Sma I
Kpn I
Sac I
EcoRI

FIG. 10

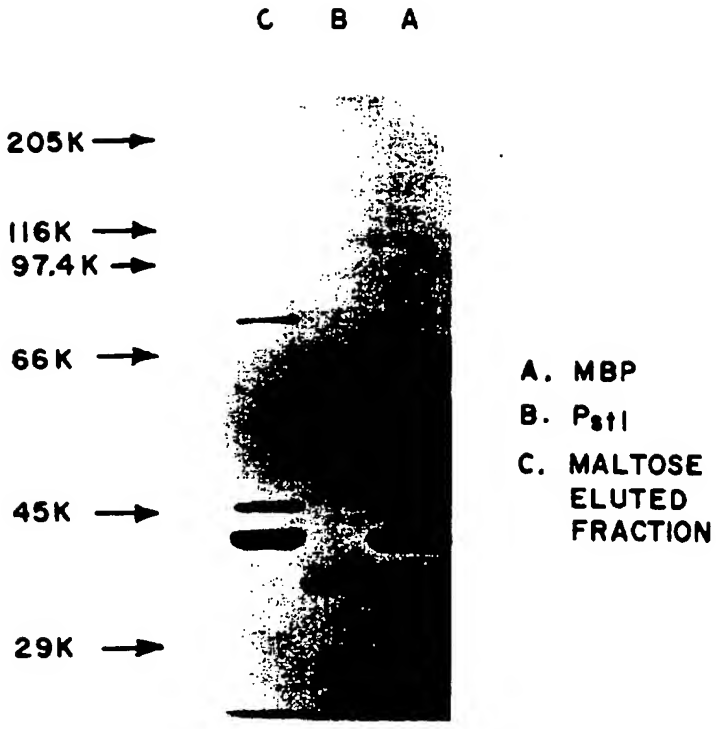

C
B
A
205K
116K
97.4K
66K
45K
29K
A. MBP
B. PstI
C. MALTOSE ELUTED FRACTION

FIG. 11

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 51, Issue of December 20, pp. 32951–32959, 1996
Printed in U.S.A.

cDNA Cloning and Sequencing Reveal the Major Horse Allergen Equ c1 to Be a Glycoprotein Member of the Lipocalin Superfamily*

(Received for publication, August 2, 1996, and in revised form, October 4, 1996)

Christophe Gregoire‡§, Isabelle Rosinski-Chupin¶, Jacques Rabillon‡, Pedro M. Alzari‖, Bernard David‡, and Jean-Pierre Dandeu‡

From ‡Unité d'Immuno-Allergie, Département de Physiopathologie, ¶Unité de Génétique et de Biochimie du Développement, Département d'Immunologie, and ‖Unité d'Immunologie Structurale, Département d'Immunologie, Institut Pasteur, 28 rue du Dr Roux, 75024 Paris Cedex 15, France

The gene encoding the major horse allergen, designated *Equus caballus* allergen 1 (Equ c1), was cloned from total cDNA of sublingual salivary glands by reverse transcription-polymerase chain reaction using synthetic degenerate oligonucleotides deduced from N-terminal and internal peptide sequences of the glycosylated hair dandruff protein. A recombinant form of the protein, with a polyhistidine tail, was expressed in *Escherichia coli* and purified by immobilized metal affinity chromatography. The recombinant protein is able to induce a passive cutaneous anaphylaxis reaction in rat, and it behaves similarly to the native Equ c1 in several immunological tests with allergic patients' IgE antibodies, mouse monoclonal antibodies, or rabbit polyclonal IgG antibodies. Amino acid sequence identity of 49–51% with rodent urinary proteins from mice and rats suggests that Equ c1 is a new member of the lipocalin superfamily of hydrophobic ligand-binding proteins that includes several other major allergens. An RNA blot analysis demonstrates the expression of mRNA Equ c1 in liver and in sublingual and submaxillary salivary glands.

Exposure to animal danders, commonly present in the environment, is known to be a frequent cause of allergy. The inhalation of these potent animal dandruff allergens induces immunoglobulin E antibody (IgE) and subsequent development of asthma in atopic individuals. Among these allergens, a major allergen is defined to be the one that elicits an anaphylactic reaction in a majority of patients, presenting an immediate hypersensitivity response mediated by IgE against the basic raw material (1).

The reasons why a protein is allergenic are not clearly understood to date, although several authors favor the hypothesis of a possible relationship between the structure and the function of proteins and their allergenicity (2). The enzymatic activity of certain proteins has been assumed to have a capacity to enhance the IgE response (2). A family of proteins, the lipocalin superfamily, is known to include several allergens, such as the mouse major urinary protein mMUP[1] (3), the rat α-2-microglobulin (rA2U) (4), the bovine β-lactoglobulin (βlg) (5), the cockroach allergen Bla g4 (6), and the recently described bovine dander allergen Bos d2 (7). Based on this observation, Arruda *et al.* suggested that lipocalins may contain a common structure that is able to induce the IgE response. Members of this superfamily, which bind or transport small hydrophobic molecules, are generally expressed in the liver and/or secretory glands. This is particularly true for the mMUP and rA2U proteins, which are multigenic families at about 35–40 members in the case of the mMUP family (8) and about 25 for the rA2U (9, 10). These members are differentially expressed in the liver as well as salivary, lachrimal, and other secretory glands (11).

The major horse allergen, Equ c1, is a potent allergen responsible for about 80% of anti-horse IgE antibody response in patients who are chronically exposed to horse allergens. Although much work has been carried out on the isolation and identification of the horse allergenic agents responsible for human hypersensitivity response (12–16), the major horse allergen was only recently purified from hair and dandruff (17). A previous study by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectric focusing-PAGE showed that Equ c1 appears as a single polypeptide with a relative molecular mass of 21,500 daltons and a pI of 3.9. The purification of Equ c1 allowed the sequencing of the 27 N-terminal amino acids and of internal peptides (18).

To obtain more information on the structural and functional features of Equ c1, we have cloned the corresponding cDNA from the sublingual salivary gland (SLG). Here we report the molecular cloning and sequencing of this cDNA and expression of a recombinant allergen rSLG Equ c1 in a bacterial system. The recombinant protein was compared with natural Equ c1 for its recognition by antibodies raised against the natural Equ c1 in immunoblots and in inhibition/competition enzyme-linked immunosorbent assay (ELISA). We also show that the recombinant protein is able to elicit a rat mast cell degranulation by passive cutaneous anaphylaxis reaction.

Sequence comparisons reveal that Equ c1 is a new member of the lipocalin superfamily.

EXPERIMENTAL PROCEDURES

Materials—The horse salivary glands were obtained from a slaughterhouse and rapidly frozen in liquid nitrogen after dissection. They were stored at −80 °C until protein and nucleotidic extractions were performed.

Protein Purification and N-terminal Sequencing—Equ c1 was puri-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U70823.

§ To whom correspondence should be addressed. Tel.: 33-1-45-68-84-48; Fax: 33-1-40-61-31-60: E-mail: cgreg@pasteur.fr.

[1] The abbreviations used are: mMUP, mouse major urinary protein; SLG, sublingual gland; HD, hair dandruff; SMG, submaxillary gland; rSLG Equ c1, recombinant SLG Equ c1; rA2U, rat α-2-microglobulin; mAb, monoclonal antibody; FPLC, fast protein liquid chromatography; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; ELISA, enzyme-linked immunosorbent assay; BSA, bovine serum albumin.

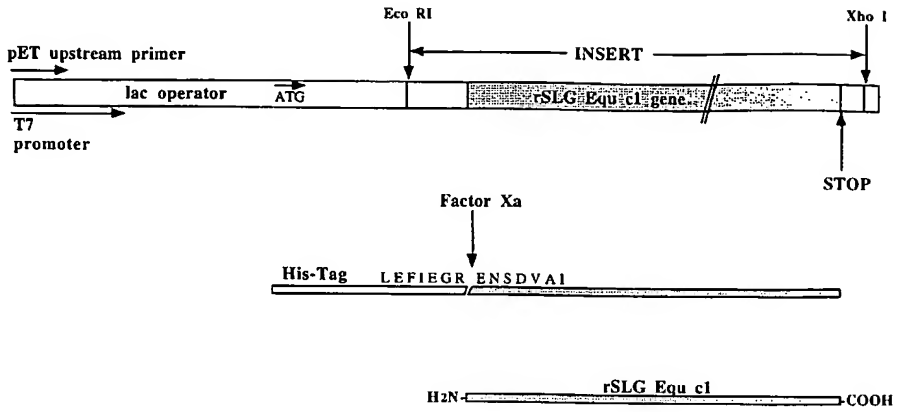


FIG. 1. **Plasmid construct for the bacterial expression of rSLG Equ c1 in *E. coli*.** cDNA Equ c1 was inserted in pET 28 (a) after digestion with *Eco*RI and *Xho*I. The plasmid contains the *lac* operator used to induce, with 1 mM of isopropyl β-D-thiogalactopyranoside, the recombinant protein tailed at its N-terminal end. Factor Xa proteolytic site (LEFIEGR↓ENSDVA) was introduced between rSLG Equ c1 ant the tail containing the polyhistidine tag.

fied from salivary glands and dander extracts by a combination of size exclusion chromatography in fast protein liquid chromatography (FPLC) and hydrophobic interaction chromatography as described previously (17).

An Equ c1 tryptic proteolysis was performed for 15 min at 37 °C in a buffer containing 50 mM Tris-HCl, 1 mM $CaCl_2$, pH 7.0, with an enzyme ratio of 1:1000 (w/w). The sequencing was processed, using the method described by Baw *et al.* (19), in the microsequencing laboratory of the Pasteur Institute. Protein assays were performed with the colorimetric method using Micro BCA protein assay reagent from Pierce, according to Smith *et al.* (20).

Preparation of RNA—Total RNA was isolated from sublingual (SLG) and submaxillary (SMG) salivary glands and from liver according to Chirgwin's protocol (21), modified as described previously (22, 23).

Equ c1 cDNA Cloning—cDNA first strand synthesis was performed on 5 μg of horse SLG total RNA for 1 h at 37 °C in a total volume of 50 μl with 20 pmol of the primer adapter oligo(dT): 5′-AAC CCG GCT CGA GCG GCC GCT TTT TTT TTT TTT TT-3′, 800 units of Moloney murine leukemia virus reverse transcriptase (Life Technologies, Inc.) in the manufacturer's buffer. The cDNAs so obtained were amplified by polymerase chain reaction (PCR) with the Opti Prime PCR optimization kit (Stratagene), with the oligomer 5′-GGY GAG TGG TAY TCY ATY TT-3′ as primer 1 and the oligomer 5′-GGY GAG TGG TAY AGY ATY TT-3′ as primer 2 derived from the Gly^{35}–Ser^{39} sequence and the 5′-GTS AGP TCR ATR ATR TTY TC-3′ as primer 3 derived from the Glu^{165}–Leu^{170} sequence. The letter Y represents a 50% mixture (w/w) of nucleotides T and C, S a mixture of G and C, and R a mixture of A and G. After a first denaturation cycle at 98 °C for 2 min, 30 cycles of PCR consisting of a 30-s denaturation step at 94 °C followed by annealing at 50 °C for 35 s and elongation at 72 °C for 30 s were carried out in a thermocycler Hybaid (Ceralabo, Aubervilliers, France). Each reaction contained 1 μl of cDNA reaction product, 0.2 mM dNTP, 2.4 units of *Taq* DNA polymerase, 68.8 pmol of the primer 3, and 34.4 pmol of each other primer. The variable parameters of buffers are pH, $MgCl_2$, and KCl concentrations. The best amplification was obtained with buffer 6 (10 mM Tris-HCl, pH 8.8, 1.5 mM $MgCl_2$, and 75 mM KCl) and buffer 12 (10 mM Tris-HCl, pH 9.2, 3.5 mM $MgCl_2$, and 75 mM KCl). After separation by electrophoresis in a 1.2% agarose gel and purification, the products from the PCR reactions were inserted in pMOS Blue T vector (Amersham Life Sciences). Sequencing was performed after alkaline denaturation by the dideoxy chain termination method (24) using Sequenase version 2.0 (U.S. Biochemical Corp.) and α-^{35}S-labeled dATP.

Amplification of the cDNA Ends—The rapid amplification of cDNA ends (RACE) strategy was applied to clone 3′ and 5′ cDNA extremities. For 5′ RACE, 12.5 μl of the first single strand cDNA (as described above) were directly used for dC tailing, for 5 min at 37 °C, in 10 mM Tris-HCl, pH 8.4, 25 mM KCl, 1.25 mM $MgCl_2$, 50 μg/ml BSA, and 10 units of terminal transferase. Reactions were stopped by increasing the temperature to 65 °C for 10 min. The cDNA amplification was performed in the presence of 5 pmol of the oligomer 5′-GCG CCC AGT GTG CTG GCT GCA GGG GGG GGG GG-3′, complementary to the dC tail, and the oligomer 5′-CTT TTC CTT GAC GTC TGA AGC C-3′ corresponding to the nucleotide sequence G^{189}–G^{210}, as a specific primer (antisense). A 5-μl aliquot of dC-tailed cDNA was amplified by PCR in a 50-μl volume in 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 2.5 mM $MgCl_2$, 100 μg/ml BSA) and 0.2 mM each dNTP. The conditions of 35 cycles of PCR consisted of a 30-s denaturation step at 95 °C followed by a 35-s annealing step at 60 °C and a 30-s extension step at 72 °C.

For cloning of the 3′ region, the same experimental conditions were applied to the PCR amplification using the specific primer 5′-GCC CGA GAA CCA GAT GTG AGT-3′ corresponding to the nucleotide sequence G^{481}–T^{501} and the primer adapter oligo(dT). All amplified products were cloned in pMOS Blue vector and sequenced as described above.

Bacterial Expression of Recombinant Equ c1—A cDNA corresponding to the nearly complete Equ c1 sequence was amplified by PCR and cloned in a pET vector. Primers for PCR were designed to specifically hybridize with Equ c1 cDNA and contained *Eco*RI and *Xho*I sites. The primers used were 5′-CTT GAA TTC ATC GAG GGG AGA GAA AAC AGT GAT GTT GCG-3′ (5′ end primer) and 5′-CCA CTC GAG GAA GTA TTC ACT GTC-3′ (3′ end primer). In addition, the 5′ primer provides the recombinant protein with a new proteolytic cleavage site for the factor Xa. PCR products were cloned into the *Eco*RI/*Xho*I sites of the plasmid pET 28 (a) under control of the T7 lac promoter (Fig. 1). This expression vector contains the kanamycin resistance gene and a His_6 tag at the N terminus of the recombinant protein. Competent *Escherichia coli* XL1 cells were transformed, and supercoiled plasmid was sequenced and transfected in *E. coli* BL 21 (DE3). Induction was performed by adding isopropyl β-D-thiogalactopyranoside to the medium at a final concentration of 1 mM for 180 min at 37 °C. Induction was controlled by taking aliquots every 30 min. Cells were then harvested by centrifugation and resuspended in 50 mM Tris-HCl, pH 7.0, containing 1% (v/v) Triton X-100 and 100 μg/ml lysozyme. The cells were incubated for 15 min at 30 °C, and the DNA was disrupted by sonication. The supernatant obtained after centrifugation was filtered on a 0.2-μm membrane and dialyzed against phosphate-buffered saline (PBS) with 0.5 mM NaCl. The resulting product was used for chromatographic purification.

Purification of the Recombinant Equ c1—An HR 5/5 column was packed with chelating Sepharose fast flow (Pharmacia Biotech, Inc.), washed according to the manufacturer's suggestions, and charged until saturation with metal ions from a 0.5% (w/v) copper(II) chloride solution. After thorough rinsing with water, the column was presaturated with buffer (PBS/0.5 mM NaCl) containing 10 mM imidazole (25). After equilibration of the column with the starting buffer (PBS/0.5 mM NaCl), 6 column volumes of supernatant was loaded, and the unbound material was collected. Competitive elution was carried out using imidazole at 40 and 120 mM (PBS/0.5 mM NaCl), pH 7.0, collecting 6 column volumes at each step (26). The whole process was controlled by an FPLC apparatus (Pharmacia). The fractions were concentrated using stirred cell ultrafiltration with a PM 10 membrane (Amicon) and dialyzed against the proteolysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM $CaCl_2$). Digestion with the factor Xa was performed overnight at 30 °C. After proteolysis, the digest was dialyzed to remove the small digest peptides and lyophilized.

SDS-PAGE and Western Blots—All analysis of the different fractions was performed with the Adjustable Stab Gel kit ASG 400 (Prolabo) using 18% acrylamide/bisacrylamide (29:1) gels (27). Proteins were visualized with Coomassie Blue and/or silver nitrate staining. Electroblotting experiments were performed using nitrocellulose membrane (Schleicher & Schüll). For immunological detection, polyclonal antibodies from human and rabbit sera and mouse monoclonal antibody directed against Equ c1 were used.

The rabbit immunization was performed by intradermal injection of 100 μg of pure allergen. Sixteen patients with established allergy to natural Equ c1 were selected, and a pool from three nonallergic healthy donors was used as negative control. Bound IgE were detected using peroxidase conjugated to rabbit anti-human IgE. When mouse mAb anti-Equ c1 or the polyclonal rabbit IgG was used, the detection was

performed with peroxidase conjugated to rabbit anti-mouse IgG or peroxidase conjugated to goat anti-rabbit IgG, respectively, using the diamino-3,3'-benzidine tetrachlorydrate as specific reagent.

The mouse anti-HD Equ c1 mAb were prepared in the Hybridolab of the Pasteur Institute according to the methods described by Köhler and Milstein (28).

Passive Cutaneous Anaphylaxis—Each mouse was immunized subcutaneously at day 0 and boosted at days 21 and 35 with 5 μg of antigen (purified HD Equ c1, protein extract from horse hair dandruff, horse serum albumin, or ovalbumin) in the presence of 4% (w/w) $Al(OH)_3$ in a physiological solution. Each mouse was bled after being anesthetized, at day 42 by retro-orbital puncture in order to study IgE immune response. The IgE antibody titers were determined by the passive cutaneous anaphylaxis reaction in rats (29).

Serum samples were diluted in a physiological solution and 100-μl aliquots inoculated intradermally on the shaved back of Lewis rats. Twenty-four hours later, each rat was challenged by intravenous inoculation in the tail of 1 ml of a physiological solution containing 50 μg of antigen and 0.5% Evans blue. Thirty minutes later, rats were killed, and skin was excised for examination. The reciprocal of the highest dilution giving a blueing reaction of 10-mm diameter was taken as the passive cutaneous anaphylaxis titer.

Inhibition/Competition Experiments—These experiments were performed using ELISA as follows. Each well of the assay plate (Maxisorb, Nunc, Roskild, Denmark) was coated with 100 μl of a highly purified HD Equ c1 or rSLG Equ c1, 10 μg/ml in 0.1 M carbonate/bicarbonate buffer, pH 9.6. After saturation of the unoccupied sites with 0.5% BSA in PBS and appropriate washing, mAbs, after being previously preincubated 1 h at 37 °C with different dilutions of competitor, were added in duplicate to the sample-coated wells and incubated for 1 h at 37 °C. Bound mAb and rabbit antibodies were detected with peroxidase-conjugated rabbit anti-mouse IgG (Sigma) and peroxidase-conjugated goat anti-rabbit IgG, respectively, and revealed with *o*-phenylenediamine according to the manufacturer's recommendations.

Determination of Sugar Content—A study was done to perform deglycosylation on Equ c1, using anhydrous trifluoromethane sulfonic acid, as described by Sojar and Bahl (30). Each dry sample was acid-treated with a mixture of trifluoromethane sulfonic acid and toluene for 4 h at −20 °C. Then trifluoromethane sulfonic acid was neutralized by adding to the reaction mixture pyridine and ammonium bicarbonate and dialyzed against 50 mM Tris/HCl, pH 7.5, 100 mM NaCl. Each sample was submitted to electrophoresis in SDS-PAGE. Gels were stained with silver nitrate. Analysis of the saccharide composition of the HD Equ c1 and Saliva Equ c1 was done using gas phase chromatography after acidic treatment, as described by Kamerling *et al.* (31).

RNA Analysis—Total mRNA was electrophoresed in an agarose/formaldehyde gel (32) transferred to a nylon membrane, and hybridized with the Equ c1 cDNA probe. The probe was the full-length cDNA insert labeled by the random priming method (33).

The search for homologies between the deduced amino acid sequence of Equ c1 and the proteins of the Swiss-Prot data base or the Equ c1 cDNA and the GenBank™ nucleotide sequence data base were done, respectively, with the FASTP and FASTN program according to Altschul *et al.* (34).

RESULTS

Molecular Cloning of the Equ c1 cDNA—Tryptic fragments were generated from HD Equ c1 isolated and purified from horse hair dandruff extract by a combination of size exclusion chromatography and hydrophobic interaction chromatography. These fragments were microsequenced, and two of them (shown in *boldface type* on Fig. 2) were used to design three degenerate primers. The design of the primers took into consideration the codon usage in horse.

It was previously demonstrated by Dandeu *et al.* (17) that Equ c1 from different sources, *i.e.* saliva, urine, and hair dandruff extracts, are similarly recognized by antibodies. Salivary secretions contain the highest amount of Equ c1 protein; therefore, the salivary glands were chosen to clone Equ c1 cDNA. Among the tested salivary glands, the sublingual glands had the highest level of Equ c1 immunoreactivity and were selected to prepare mRNA.

The mRNAs so obtained were reverse-transcribed, and the Equ c1 cDNA was amplified by PCR using a mixture of the three primers. This reverse transcription-PCR resulted in a DNA fragment of about 400 base pairs in length that was cloned in pMOS Blue; several positive clones were sequenced. In a second step, 5′ and 3′ ends of the SLG Equ c1 cDNA were obtained using a 5′ and 3′ RACE strategy. The two amplification products of 250 and 450 base pairs for the 5′ and 3′ RACE, respectively, were cloned and sequenced.

Sequence of the Equ c1 cDNA—The full-length sequence of Equ c1 cDNA and the deduced amino acid sequence are shown in Fig. 2. The SLG Equ c1 cDNA is 923 nucleotides long with an open reading frame of 560 nucleotides (excluding the stop codon), coding for a 187-amino acid protein. All peptides from HD Equ c1 can be localized in the SLG Equ c1 sequence and start after an arginine or a lysine residue, according to the tryptic proteolysis consensus sites. However, some differences in the amino acid sequence can be observed between rSLG Equ c1 from sublingual salivary gland and the tryptic peptides obtained from HD Equ c1. These differences are not PCR artifacts, because our nucleotide sequence results from the analysis of 12 clones from four independent PCR experiments. These differences are in the internal peptides, at positions 62 (Ala/Leu), 90 (Phe/Ala), 136 (Phe/Leu), 146 (Ser/Asp), 172 (Lys/Gln), and 173 (Ile/Thr). All analyzed clones contained a 3′ noncoding region of 298 nucleotides and a poly(A) tail 23 base pairs, downstream from a consensus polyadenylation signal AATAAA at position A^{886}/A^{891}. All clones sequenced have identical 5′ ends with a noncoding region of 63 nucleotides and with an open reading frame beginning at A^{64}.

Analysis of the deduced amino acid sequence revealed that the 5′ end of the coding region contains a typical signal sequence (35) (Fig. 3*A*). According to the Von Heijne weight matrix method (36), a favored putative signal peptidase cleavage site can be assigned between the Ala^{15} and Gln^{16} residues, generating a protein beginning with QQEENSDVAI. In contrast, the N-terminal end of the protein initially purified from hair dandruff (SDVAI) would result from a cleavage between Asn^{20} and Ser^{21}, which is not predicted by Von Heijne's rules. Equ c1 was purified from saliva, and the microsequencing of its N-terminal peptide revealed a mixture of three sequences, one of them beginning at the predicted Gln^{16}, but the others at Glu^{18} and Ser^{21}, respectively (Fig. 3*B*). Whether these N-terminal ends are due to cleavage by signal peptidase at different sites or are generated by proteolytic processing of the secreted protein is not known. Such heterogeneous N-terminal ends were also reported for human tear albumin (37), another member of the lipocalin superfamily

Excluding the putative signal peptide, the protein contains two cysteine residues at positions 83 and 176. In a previous study, we observed an increase in the apparent molecular mass of Equ c1 from 21,500 to 25,000 daltons in SDS-PAGE gels under reducing conditions, indicating that these two cysteines could form a disulfide bridge. Equ c1 is highly rich in charged residues and aromatic residues. The calculated pI is 4.57, a value close to that determined by Dandeu *et al.*

Glycosylation of Equ c1—Two putative *N*-glycosylation sites are present at positions Asn^{53} and Asn^{68}. Glycosylation of HD and SLG Equ c1 was confirmed by gas phase chromatography, which revealed the presence of approximately 8.6% (w/w) of carbohydrates, representing 1,850 daltons. These results could explain the decrease in apparent molecular weight of Equ c1 in SDS-PAGE (Fig. 4) and the modification of the pI after deglycosylation.

Analysis of the sugar residue composition in Table I shows the presence of GalNAc, Gal, NeuAc, GlcNAc, and Man. Carbohydrates attached to proteins can be classified into two groups, *N*-glycans and *O*-glycans. All *N*-glycans contain a common structure, Manα1→6(Manα1→3)Manβ1→4GlcNAcβ1→

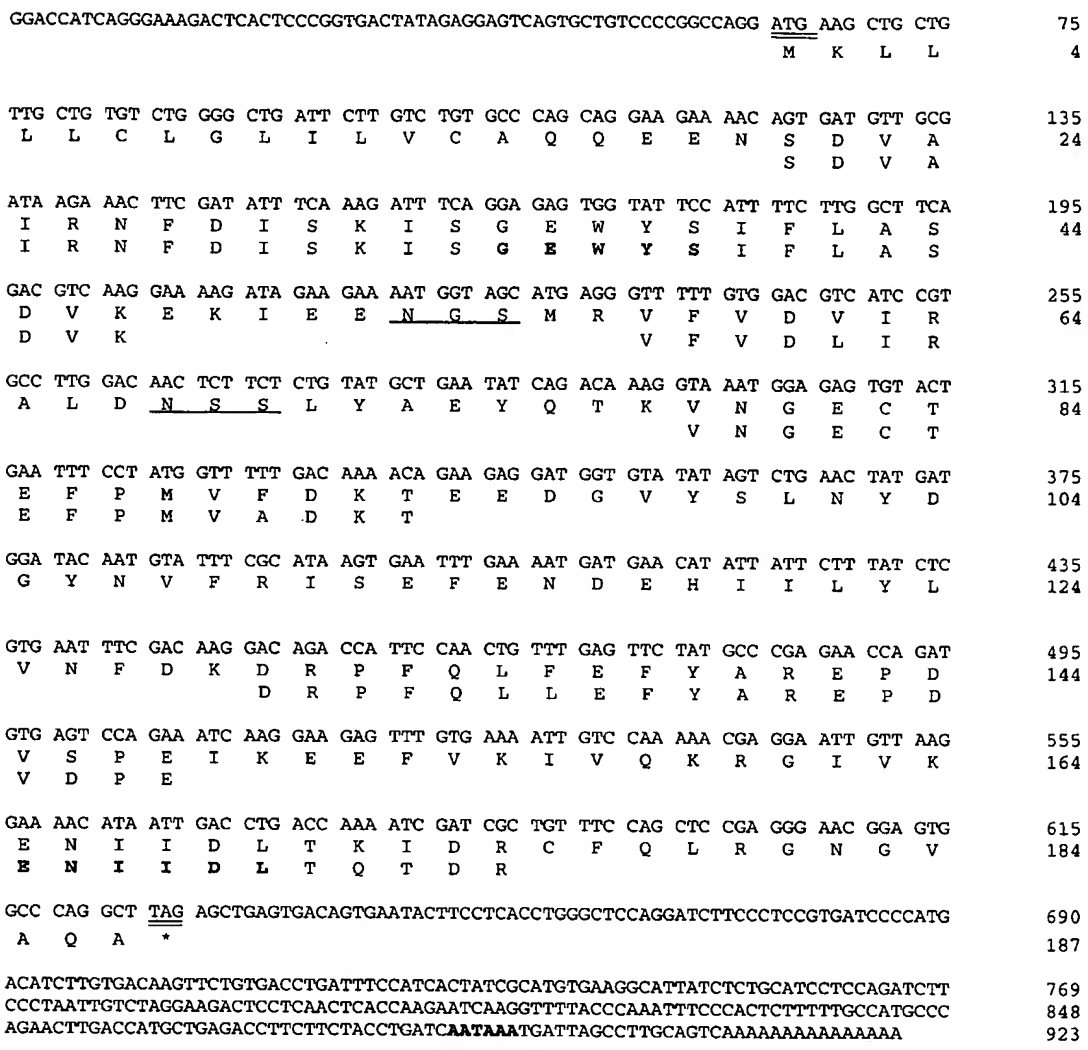


FIG. 2. **Full-length sequence of the SLG Equ c1 cDNA (*first line*) and its deduced amino acid sequence (*second line*).** The N-terminal and internal tryptic fragments of HD Equ c1 are positioned in the *third line*. The start and stop codons are *double underlined*, and the polyadenylation signal is in *boldface type*. The *N*-glycosylation sites are *underlined*, and the amino acid sequences used to deduce the sequence for the degenerate primers are in *boldface type*.

4GlcNAc→Asn, called the trimannosyl core. Molecular ratio results (second column in Table I) indicate unambiguously the presence of this core and, therefore, the presence in the glucidic part of Equ c1 of one *N*-glycan member of the biantennary complex type that contains three mannose residues. One *N*-acetyl-lactosamine (Galβ1→4GlcNAc) is attached to the outer two α mannose residues, followed by sialic acid residues (for Equ c1) or additional *N*-acetyllactosamines (38).

The presence of GalNAc only found in the *O*-glycan components, except for several hormones (38), suggests that the protein is also *O*-glycosylated.

Expression of Equ c1 as a Recombinant Protein—A recombinant protein, starting at Glu^{19}, was produced in a bacterial system, after cloning of the corresponding cDNA sequence in a pET 28 plasmid. This plasmid allows bacterial expression of a recombinant protein with a 40-amino acid polypeptide tail containing a polyhistidine tag to its N-terminal end (Fig. 1). To allow the production of a recombinant protein without any added amino acid, a factor Xa proteolytic site (LEFIEGR↓ENSDVA) was inserted between the tail and the recombinant protein.

Two recombinant clones were tested for rSLG Equ c1 expression. Optimal production was obtained after a 150-min induction by isopropyl β-D-thiogalactopyranoside. A protein determination assay showed that rSLG Equ c1 represents about 30% of the total bacterial protein. This protein was essentially present in the supernatant of the bacterial extracts. A single purification step by immobilized metal affinity chromatography was sufficient to obtain pure rSLG Equ c1, which migrates as a single band of 19.5–20 kDa in an 18% SDS-PAGE gel (Fig. 4, *lane A*) after cleavage by factor Xa. This molecular mass is compatible with the calculated mass of 19,469 daltons and rather similar to that of deglycosylated natural Equ c1.

Antigenicity of the Recombinant Protein—The recombinant protein was tested for its antigenic recognition by different antibodies raised against HD Equ c1, *i.e.* three mouse monoclonal antibodies (mAbs 118 and 197, which recognize two different linear epitopes, and mAb 220, which recognizes a conformational epitope),[2] mouse and rabbit polyclonal antibodies (IgG), and human IgE from the sera of 16 patients suffering from horse allergic reactions (characterized in Ref. 17).

Immunoblot analysis after SDS-PAGE (Fig. 5), performed on the total bacterial extract, shows that the three mAbs bind a

[2] C. Gregoire, J. Rabillon, B. David, and J.-P. Dandeu, unpublished data.

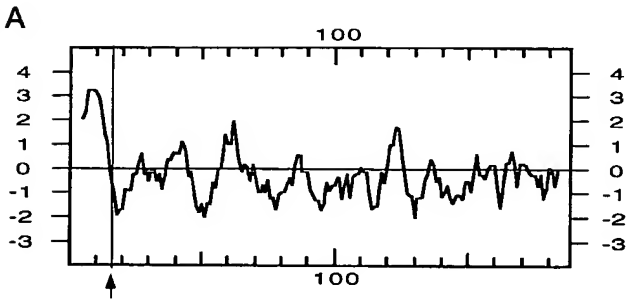


FIG. 3. **Hydrophobicity profile of the SLG Equ c1 calculated by the Kyte and Doolittle method (53).** *A*, the putative favored peptide signal cleavage predicted by Von Heijne's rule (36), between Ala^{15} and Gln^{16} is shown. *B*, N-terminal amino acid sequences obtained by direct sequencing of the major horse allergen Equ c1 purified from salivary extract (*lanes 1–3*) and from horse's hair dandruff extract (*lane 4*) according the microsequencing method described by Bauw *et al.* (19).

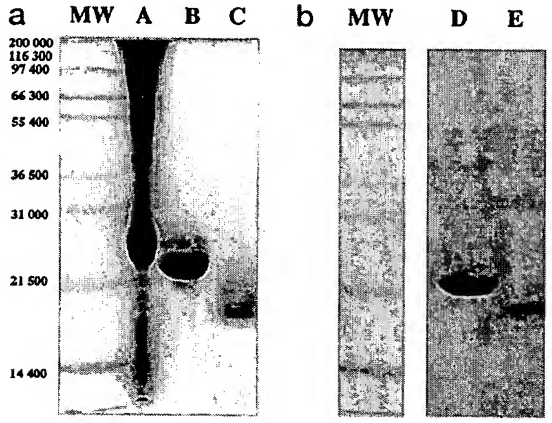


FIG. 4. **SDS-PAGE in 18% acrylamide/bisacrylamide (29:1) gel of the rSLG Equ c1, expressed in *E. coli* as a polyhistidine-tailed protein.** *a*, 10-μl sample of total bacterial extract was electrophoresed (*lane a*), followed by purification by immobilized metal affinity chromatography (*lane B*) and cleavage with factor Xa (*lane C*). Proteins were visualized by gel Coomassie Blue staining. *b*, 1 μg of HD Equ c1 was electrophoresed (*lane D*), followed by 1 μg of total deglycosylated HD Equ c1 (*lane E*). The gel was stained with silver nitrate. The molecular weight markers visualized by Coomassie blue staining are indicated.

24-kDa single band corresponding to the recombinant protein with the His tag. The tailed rSLG Equ c1 is also recognized by polyclonal anti HD Equ c1 antibodies from mouse and rabbit sera, although the latter also binds a contaminating band around 36 kDa. In contrast, rSLG Equ c1 is not recognized by rabbit or mouse control sera from animals immunized with horse serum albumin or ovalbumin.

In addition, rSLG Equ c1 is also recognized by the sera of allergic patients in Western blot experiments, suggesting that some or all of the HD Equ c1 epitopes recognized by human IgE are also present on the rSLG Equ c1. Fifteen other sera of allergic patients with established allergy to natural Equ c1 were tested. The same results were obtained with all of these antisera (data not shown). Sera from nonallergic patients failed to detect rSLG Equ c1.

Inhibition/competition experiments with the three different mAbs in an ELISA were performed using rSLG Equ c1, after purification and proteolysis by the factor Xa, and using pure

TABLE I
Determination of monosaccharide composition

The sugar content was determined by gas phase chromatography on pure HD/SLG Equ c1 (31). The relative weight ratio was given for each monosaccharide. The molecular ratio (column 2) was compared with the theoretical ratio for one *N*-glycan biantennary complex type given in parenthesis.

Monosaccharide	Weight	Molecular ratio
	%	
Man	1.8	3.0 (3)
Gal	1.5	2.5 (2)
GalNAc	0.53	0.7 (0)
GlcNac	2.6	3.5 (4)
NeuAc	2.2	2.5 (2)

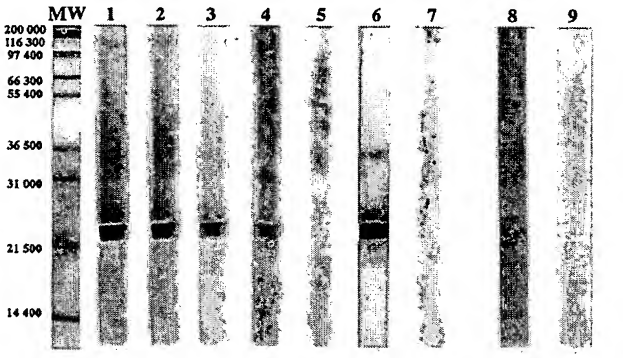


FIG. 5. **Immunoblot analysis of total bacterial extract after SDS-PAGE analysis.** The electrophoresis and blotting was performed using 10 μl of bacterial supernatant (diluted 1:500). Immunological detection was performed using monoclonal antibodies mAb 118 (*lane 1*), mAb 197 (*lane 2*), mAb 220 (*lane 3*), a polyclonal mouse serum (*lane 4*), and polyclonal rabbit serum (*lane 6*) raised against HD Equ c1 and using IgE from human serum (representative of 16 tested human sera) of a patient suffering from allergic reactions to horses (*lane 8*). Nonimmunized mouse (*lane 5*) and rabbit (*lane 7*) sera and human serum from nonallergic patients (*lane 9*) were used as controls.

HD Equ c1. The results in Fig. 6*A* show that preincubation of mAb 220 with an adequate rSLG Equ c1 or HD Equ c1 concentration completely abolished its binding to natural HD Equ c1 coated on the plates. The IC_{50} (concentration of inhibitor giving a 50% inhibition) was obtained with the same concentration of rSLG Equ c1 and of HD Equ c1, approximately 100 ng/ml. Similar results were obtained when the plates were coated with the rSLG Equ c1 protein. Experiments using the two other mAbs reveal that rSLG Equ c1 and HD Equ c1 are similarly recognized (data not shown). No competition was observed when BSA was used as a competitor.

The inhibition/competition experiment performed with the polyclonal antibodies from rabbit sera raised against HD Equ c1 (Fig. 6*B*) reveals similar competition profiles when rSLG Equ c1 or HD Equ c1 are used as competitors; 100 and 50% inhibition are obtained with 20 μg/ml and 100 ng/ml, respectively, of either of them. This result suggests that the majority of the HD Equ c1 epitopes are present on the recombinant protein structure.

The biological activity of rSLG Equ c1 was also tested by passive cutaneous anaphylaxis on several rats as described under "Experimental Procedures." The mouse sera were harvested after animal immunization with HD Equ c1, hair dandruff extract, or control proteins (horse serum albumin or ovalbumin). The results in Table II show that rSLG Equ c1 elicits a positive reaction with the mouse anti-HD Equ c1 and the anti-horse hair dandruff sera. These positive reactions are obtained with rSLG Equ c1 and with HD Equ c1 at the same serum dilution. In the same conditions rSLG Equ c1 did not display any positive reaction with the control sera.

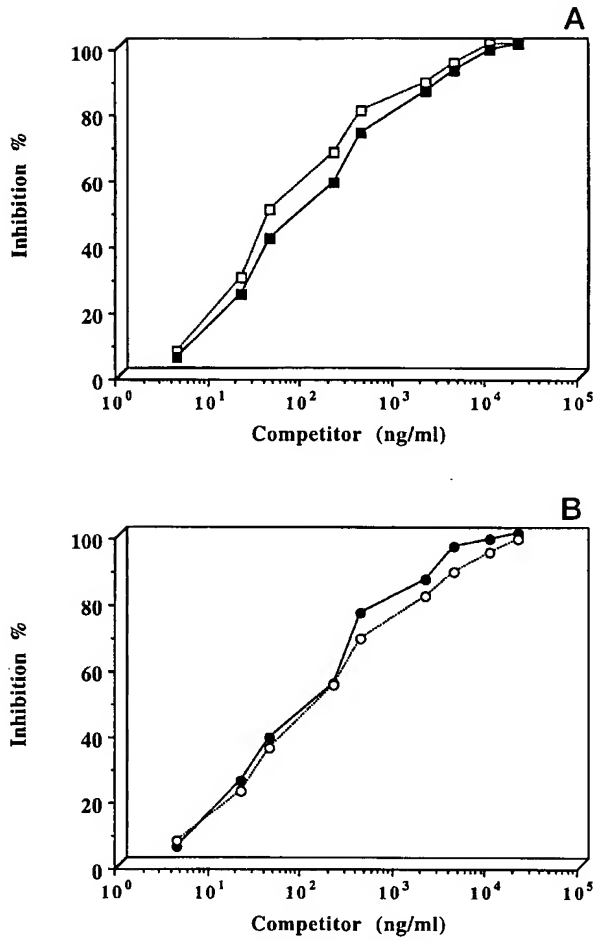


FIG. 6. **Competition in ELISAs.** Plates were coated with 100 μl of HD Equ c1 (100 ng/ml) in 0.1 M carbonate/bicarbonate buffer. Competition was performed by preincubation of mAb 220 (*A*) or polyclonal antibodies from rabbit sera raised against HD Equ c1 (*B*) for 1 h at 37 °C with 100 μl of continuous dilution of rSLG Equ c1 (*open symbols*) or HD Equ c1 (*filled symbols*) in PBS. The percentage of inhibition was calculated according to the following expression.

$$m\%\ \text{inhibition} = \frac{((\text{OD without competitor}) - (\text{OD with competitor}))}{(\text{OD without competitor})} \times 100 \quad \text{(Eq. 1)}$$

Homologies of Equ c1 with Proteins of the Lipocalin Superfamily—Homology searches in the sequence data bases show that Equ c1 has sequence similarities with other members of the lipocalin superfamily (Fig. 7). The best score was obtained with the mouse major urinary proteins cLac1 MUP4, the cSmx1 MUP5 (cloned from lachrimal and submaxillary glands, respectively), and rA2U with homology ranging from 49 to 51% of identity and 76% of conservative mutations.

Sequence alignment shows that the two cysteines, Cys^{83} and Cys^{176}, that form a disulfide bond in mMUP and rA2U, as well as the majority of other lipocalins, are conserved (39). Only one potential *N*-glycosylation site, corresponding to position Asn^{53}, is present in rA2U and is absent from the mMUP. The other site, at position Asn^{68}, is specific to Equ c1 and is due to the insertion of a serine residue at position 69.

Three motifs, relatively well conserved among lipocalins, have been described by Flower *et al.* (40). Two of these motifs are found in Equ c1 (Fig. 7). The most highly conserved amino acid sequences with the lipocalin superfamily are Lys^{32}, **Gly^{35}-*Xaa*^{36}-Trp^{37}-Tyr^{38}**, Ile^{40}, Leu^{42}-Ala^{43}-Ser^{44}-Asp^{45} in motif 1

TABLE II
PCA titers of presensitized mice

Each BALB/c mouse was immunized at day 0 with 5 μg of antigen in physiological solution with 4% (w/w) of $Al(OH)_3$ and boosted at days 21 and 35. Each mouse serum was collected at day 42. 100 μl of each of the dilutions of mouse serum was inoculated intradermally in the shaved back of a Lewis rat, and 24 h later the challenge was performed by intravenous inoculation in the rat tail of 50 μg of antigen. The PCA titers reported in the table are the highest dilution of mouse sera giving mast cell degranulation.

Mouse sera	Nature of the protein challenger				
	HD Equ c1	rSLG Equ c1	HoSA[a]	Ovalbumin	*E. coli* extract
	-fold dilution				
Anti-HD Equ c1	1280	1280	0	0	0
Anti-HoSA	0	0	640	0	0
Anti-ovalbumin	0	0	0	810	0
Control serum	0	0	0	0	0

[a] HoSA, horse serum albumin.

and **Arg^{141}-Glu^{142}-Pro^{143}-Asp^{144}**, Ile^{149}-Lys^{150}-Glu^{151}, Phe^{153} in motif 3 (41). The other conserved motif is TDY (structurally conserved region 2), while Phe^{109}, Ile^{111}, and Asp^{117} seem to be less conserved in the Equ c1 sequence. However, this motif is also absent from a number of true lipocalin members, such as the human tear albumin (37), von Ebner's gland protein (42), and hamster aphrodisin (43), and is less conserved in the bilin-binding protein (44), the α1-microglobulin (45), and rat odorant protein (46).

Tissue Expression of Equ c1 mRNA—To study the distribution of Equ c1 in the horse, total RNA was prepared from SLG and SMG salivary glands as well as from the liver, and it was analyzed by RNA blot hybridization (Fig. 8). Equ c1 mRNA was detected in each twice; however, the level in the SMG and liver is about 100 times lower than in the SLG. In addition, Equ c1 mRNA in liver seems to be slightly longer. Whether this is due to a true difference of size or to the presence of a longer poly(A) tail in liver Equ c1 mRNA was not investigated.

DISCUSSION

This paper reports the cloning, characterization, and expression in a bacterial system of the cDNA corresponding to a major horse allergen, Equ c1. This cDNA was cloned from the SLGs and some differences were noted between its deduced amino acid sequence and peptides generated from a protein purified from horse hair dandruff extract (HD Equ c1). Indeed, 6 amino acids out of 79 are different between the two sequences. Some of these changes are conservative. One likely explanation of these differences is that HD Equ c1 and SLG Equ c1 belong to the same multigenic family, whose members are tissue-specifically expressed, as was reported for rodent urinary proteins from mouse and rat (47). During the cloning of SLG Equ c1, we obtained no evidence for another member of this family being expressed in salivary sublingual glands; however, we cannot exclude the possibility that the choice of primers for reverse transcription-PCR might have favored the cloning of one cDNA only. An RNA blot study revealed the presence of mRNAs hybridizing with SLG Equ c1 cDNA in submaxillary glands and in liver too. Synthesis in the liver could explain the presence of Equ c1 in the horse's urine (18), since it was reported for proteins of the MUP family in rat and mouse (48).

Despite the slight differences in their amino acid sequences and the absence of glycosylation in rSLG Equ c1, rSLG Equ c1 and HD Equ c1 are similarly recognized in our immunoblotting studies and inhibition/competition ELISA experiments. Moreover, the results obtained in inhibition/competition ELISA with three mAbs and with rabbit antibodies raised against HD Equ c1 suggest that all IgG epitopes of HD Equ c1 are also present in rSLG Equ c1, and thus in SLG Equ c1. In addition, at least

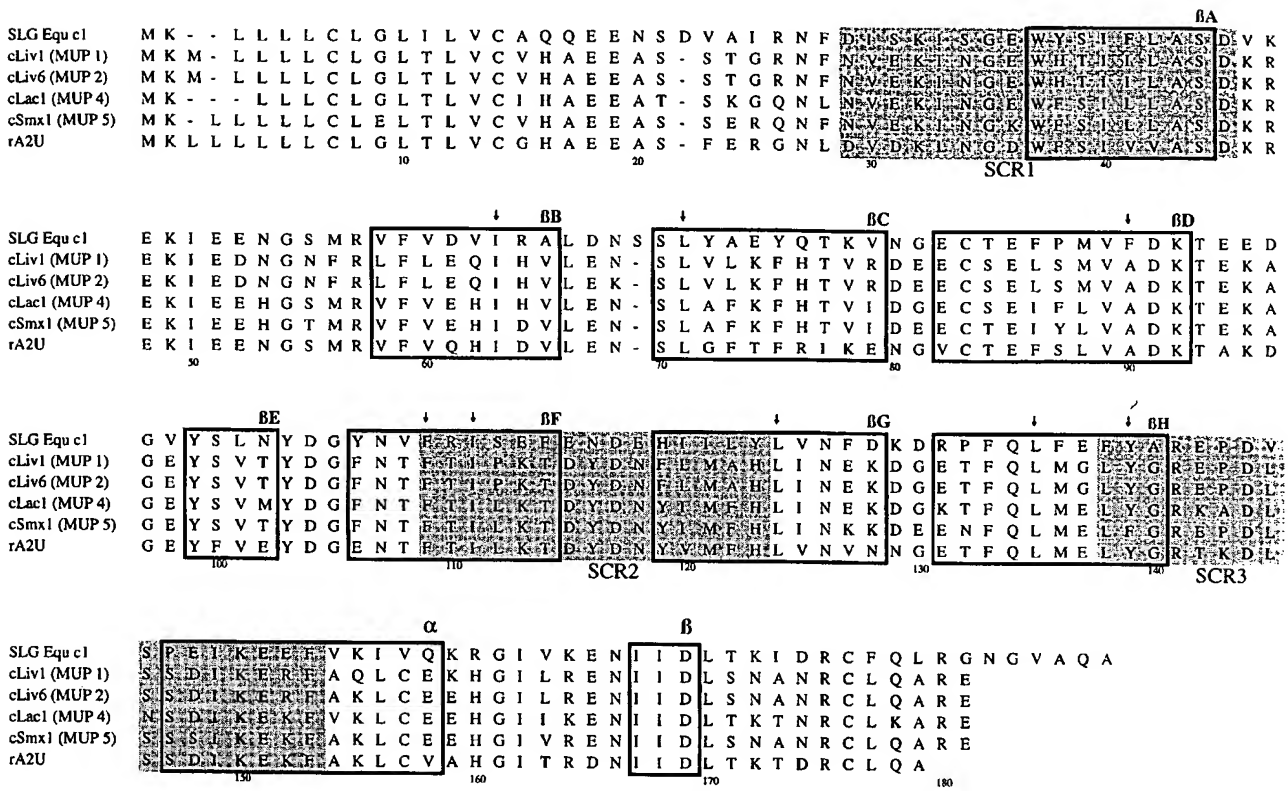


FIG. 7. **Sequence alignment of SLG Equ c1 with lipocalins.** The structurally conserved regions (*SCR1*, *SCR2*, and *SCR3*) described by Flower *et al.* (40) are shown in *gray*. Secondary structure elements from the crystal structure of MUP1 (Protein Data Bank code 1 MUP; Bocskei *et al.* (51)), as defined by the computer program DSSP (54) are *boxed*. Amino acid residues forming the binding pocket are indicated by *arrows*.

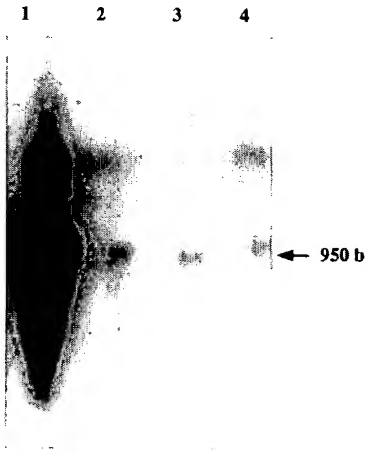


FIG. 8. **Tissue distribution of Equ c1 in horses.** Twenty micrograms and 0.2 μg of total RNA from sublingual salivary glands (*lanes 1* and *2*, respectively), 20 μg of submaxillary salivary glands (*lane 3*), and 20 μg of liver (*lane 4*) were electrophoresed in a 2% agarose/formaldehyde gel, blotted, and hybridized with the Equ c1 cDNA probe. The length of Equ c1 was estimated to be 950 bases as indicated.

some of the IgE epitopes are also present in rSLG Equ c1, since rSLG Equ c1 is recognized by IgE from allergic patients in immunoblot experiments and binds to mouse IgE in passive cutaneous anaphylaxis experiments, resulting in the induction of a specific immediate hypersensitivity response in rats presensitized with HD Equ c1. Together, these results suggest that neither the differences in amino acids nor the absence of glycosylation in the bacterially expressed protein affects the global conformation of the protein.

The search in the sequence data base revealed homology with members of the lipocalin superfamily, in particular with cLac1 MUP4 and cSmx1 MUP5. Members of this family share a common structure as was shown by the x-ray crystal structures of retinol-binding protein (49), β-lactoglobulin (50), and MUP (51). The folding architecture of lipocalins consists of an eight-stranded β-barrel followed by a single α-helix and a short C-terminal β-strand (Fig. 9). The eight anti-parallel strands are arranged in two orthogonal β-sheets that leave a small hydrophobic cavity within the barrel (52). This pocket is in a highly apolar environment, appropriate for binding and transport of small hydrophobic molecules through a hydrophilic media. The binding pocket is entirely formed by aliphatic and aromatic side chains from the inner faces of the two β-sheets (these positions are indicated by *arrows* in the alignment shown in Fig. 7).

A structural model of Equ c1 (Fig. 9) was constructed from the x-ray coordinates of the mouse MUP1 model by Böcskei *et al.* (51) using the program QUANTA (MSI). This modeling was facilitated by the absence of amino acid insertions and deletions between the two proteins, with two exceptions: the insertion of Asp^{22} at the N terminus and Ser^{69} in the β-hairpin loop between the second and the third strands of Equ c1. At positions where the two proteins differed, the amino acid sequence was substituted, and the side chains were rebuilt using stereochemical criteria. The model was finally submitted to an overall energy minimization. As can be seen in Fig. 9, many of the amino acids of the presumed binding pocket (Ile^{63}, Leu^{71}, Phe^{109}, Ile^{111}, Leu^{124}, Leu^{135}, and Tyr^{139}) are either strictly conserved or have conservative amino acid substitutions in SLG Equ c1 when compared with rA2U/mMUP. The most noticeable differences are the substitution of Ala^{73} in Equ c1 by Leu/Phe in rA2U/mMUP and the substitution of Phe^{90} in the adjacent β-strand of Equ c1 by alanine. Although the hydrophobic character of the binding pocket is maintained, these

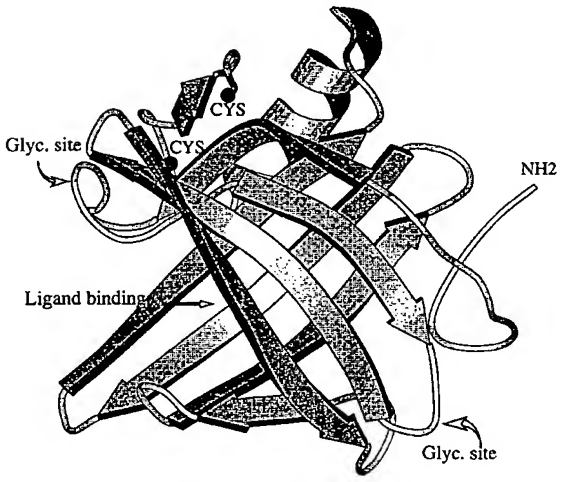


FIG. 9. **Molecular model of SLG Equ c1.** Schematic view of the lipocalin fold. The positions of the cysteine residues (*small circles*), putative *N*-glycosylation sites in SLG Equ c1, and the entrance of the binding pocket are indicated.

changes might modulate its shape and specificity.

In addition, the two possible *N*-glycosylation sites, which are not present in MUP1, are found in Equ c1 in exposed protein loops accessible to the solvent (Fig. 9), suggesting that the presence of an *N*-glycan does not interfere with the structure of the binding pocket. Moreover, the two cysteine residues that form a disulfide bridge linking the C-terminal part of the protein to the β-barrel (Fig. 9) in rA2U/mMUP (Fig. 7) and in the majority of other lipocalins are also conserved in Equ c1 (positions 83 and 176).

This structural model, therefore, suggests that Equ c1 could adopt the same tertiary structure as that described for other lipocalins. The exact physiological role of Equ c1 has not been established yet. Its presence in the urine of adult mares and stallions and its absence in the urine of yearlings (18) suggests that Equ c1 is only synthesized at sexual maturity. Thus, its physiological role could be similar to that of rodent urinary protein of mice and rats (pheromone-binding protein) but not completely identical, since these two proteins are essentially produced in males.

Our results allow us to add Equ c1 to the list of lipocalins able to induce an IgE response, thus enhancing the hypothesis of Arruda (6) that lipocalins could have an intrinsic property to stimulate the IgE production. The reasons why some members of the lipocalin superfamily are allergenic are not clear to date. One reason could be their high concentration in secretion in contact with humans, facilitating the captivation of these allergens. Indeed, Equ c1 is highly concentrated in secretory fluid such as saliva and urine as well as in hair dandruff extract (17). In addition, lipocalins have a highly conserved structure that confers a resistance to degradation. For example, βlg is able to resist acidic treatment and to pass the stomach intact (5). It has been suggested that this resistance may be important for immunogenicity.

Alternatively, there could be a link between the allergenicity of lipocalins and their small hydrophobic ligand transport function. However, such a link has not yet been established. In fact, the nature of the binding ligand differs between the lipocalins (retinol for βlg and several different pheromones for MUP and rA2U). The exact nature of the binding molecule is not known for a number of them such as Bos d2, Bla g4, and Equ c1. Last, we cannot exclude the possibility that, because of their sequence and structure similarities, lipocalins may share common epitopes important for IgE recognition. However, the existence of such a cross-reactivity remains to be clearly established.

In this context, where some members of the lipocalin superfamily may have an intrinsic property to stimulate IgE production, the obtainment of a recombinant wild-type protein and of suitable mutants that can induce a biological activity will be an important tool to study the determinants involved in allergic reactions. Morever, rSLG Equ c1 may also help in the diagnosis of the allergic reaction to horses.

Acknowledgments—We thank Prof. F. Rougeon for continuous support and useful discussions and Dr. T. Fontaine for the determination of the monosaccharide composition. We also thank Dr. Bernadac for collection of horse saliva and hair dandruff extracts, and Dr. B. Laoide for a critical reading of this manuscript.

REFERENCES

1. Lowenstein, H., Markussen, B., and Weeke, B. (1976) *Int. Arch. Allergy Appl. Immunol.* **51,** 38–47
2. Dudler, T., Machado, D. C., Kolbe, L., Annand, R. R., Rhodes, N., Gelb, M. H., Koelsch, K., Suter, M., and Helm, B. A. (1995) *J. Immunol.* **155,** 2605–2613
3. Lorusso, J. R., Moffat, S., and Ohman, J. L. (1986) *J. Allergy Clin. Immunol.* **78,** 928–937
4. Walls, A. F., and Longbottom, J. L. (1985) *J. Allergy Clin. Immunol.* **75,** 242–251
5. Ball, G., Shelton, M. J., Walsh, B. J., Hill, D. J., Hosking, C. S., and Howden, M. E. (1994) *Clin. Exp. Allergy* **24,** 758–764
6. Arruda, L. K., Vailes, L. D., Hayden, M. L., Benjamin, D. C., and Chapman, M. D. (1995) *J. Biol. Chem.* **270,** 31196–31201
7. Mantyjarvi, R. A., Rytkonen, M., Pentikainen, J., Rautiainen, J., Virtanen, T., Santa, H., and Laatikainen, R. (1996) *J. Allergy Clin. Immunol.* **97,** 212 (Abstr. 117)
8. Shahan, K., Gilmartin, M., and Derman, E. (1987) *Mol. Cell. Biol.* **7,** 1938–1946
9. Unterman, R. D., Lynch, K. R., Nakhasi, H. L., Dolan, K. P., Hamilton, J. W., Cohn, D. V., and Feigelson, P. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 3478–3482
10. Dolan, K. P., Unterman, R., McLaughlin, M., Nakhasi, H. L., Lynch, K. R., and Feigelson, P. (1982) *J. Biol. Chem.* **257,** 13527–13534
11. Gao, F., Endo, H., and Yamamoto, M. (1989) *Nucleic Acids Res.* **17,** 4629–4636
12. Stanworth, D. R. (1957) *J. Biochem.* **65,** 582–605
13. Ceska, M., and Hulten, E. (1972) *Int. Arch. Allergy Appl. Immunol.* **43,** 427–433
14. Ponterius, G., Brandt, R., Hulten, E., and Yman, L. (1973) *Int. Arch. Allergy Appl. Immunol.* **44,** 679–691
15. Lowenstein, H. (1978) *Int. Arch. Allergy Appl. Immunol.* **57,** 349–357
16. Franke, D., Maasch, H. J., Wahl, R., Schultze-Werninghaus, G., and Bretting, H. (1990) *Int. Arch. Allergy Appl. Immunol.* **92,** 309–317
17. Dandeu, J. P., Rabillon, J., Divanovic, A., Carmi-Leroy, A., and David, B. (1993) *J. Chromatogr.* **621,** 23–31
18. Dandeu, J. P., Rabillon, J., Carmi-Leroy, A., Divanovic, A., Carmoin, L., and David, B. (1995) *J. Allergy Clin. Immunol.* **95,** 348 (Abstr. 830)
19. Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B., and Celis, J. E. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7701–7705
20. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Anal. Biochem.* **150,** 76–85
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
22. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* **107,** 303–314
23. Tronik, D., Dreyfus, M., Babinet, C., and Rougeon, F. (1987) *EMBO J.* **6,** 983–987
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
25. Sulkowski, E. (1985) *Trends Biotechnol.* **3,** 1–7
26. Casey, J. L., Keep, P. A., Chester, K. A., Robson, L., Hawkins, R. E., and Begent, R. H. (1995) *J. Immunol. Methods* **179,** 105–116
27. Laemmli, U. K. (1970) *Nature* **227,** 680–685
28. Kohler, G., and Milstein, C. (1975) *Nature* **256,** 495–497
29. Ovary, Z., Caiazza, S. S., and Kojima, S. (1975) *Int. Arch. Allergy Appl. Immunol.* **48,** 16–21
30. Sojar, H. T., and Bahl, O. P. (1987) *Methods Enzymol.* **138,** 341–350
31. Kamerling, J. P., Gerwig, G. J., Vliegenthart, J. F., and Clamp, J. R. (1975) *Biochem. J.* **151,** 491–495
32. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 5201–5205
33. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
34. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
35. Von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105
36. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690
37. Redl, B., Holzfeind, P., and Lottspeich, F. (1992) *J. Biol. Chem.* **267,** 20282–20287
38. Baenziger, J. U., and Green, E. D. (1988) *Biochim. Biophys. Acta* **947,** 287–306
39. Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990) *Proteins* **8,** 44–61
40. Flower, D. R., North, A. C., and Attwood, T. K. (1993) *Protein Sci.* **2,** 753–761
41. Nagata, A., Suzuki, Y., Igarashi, M., Eguchi, N., Toh, H., Urade, Y. and Hayaishi, O. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 4020–4024

42. Schmale, H., Holtgreve-Grez, H., and Christiansen, H. (1990) *Nature* **343**, 366–369
43. Henzel, W. J., Rodriguez, H., Singer, A. G., Stults, J. T., Macrides, F., Agosta, W. C., and Niall, H. (1988) *J. Biol. Chem.* **263**, 16682–16687
44. Suter, F., Kayser, H., and Zuber, H. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 497–505
45. Kaumeyer, J. F., Polazzi, J. O., and Kotick, M. P. (1986) *Nucleic Acids Res.* **14**, 7839–7850
46. Pevsner, J., Reed, R. R., Feinstein, P. G., and Snyder, S. H. (1988) *Science* **241**, 336–339
47. MacInnes, J. I., Nozik, E. S., and Kurtz, D. T. (1986) *Mol. Cell. Biol.* **6**, 3563–3567
48. Shahan, K., Denaro, M., Gilmartin, M., Shi, Y., and Derman, E. (1987) *Mol. Cell. Biol.* **7**, 1947–1954
49. Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P. A. (1984) *EMBO J.* **3**, 1451–1454
50. Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C., Findlay, J. B., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E., and Kraulis, P. J. (1986) *Nature* **324**, 383–385
51. Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavaggioni, A., Findlay, J. B., and North, A. C. (1992) *Nature* **360**, 186–188
52. North, A. C. (1989) *J. Mol. Graph.* **7**, 67–70
53. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
54. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637